MODELING AND RECONSTRUCTION ALGORITHMS FOR DETECTION, LOCATION, AND IDENTIFICATION OF SUBSURFACE ANOMALIES

by

Constantine A. Balanis and Mark Frank

Final Report

June 1, 1990

Prepared by

Department of Electrical Engineering
Telecommunications Research Center
Arizona State University
Tempe, AZ 85287-7206

Sponsored by
Geosciences Division
U.S. Army Research Office
Research Triangle Park, NC 27709-2211
Under Contract No. DAAG29-85-K-0045

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE

3. REPORT TYPE AND DATES COVERED
Final 15 Apr 85 - 14 Apr 90

4. TITLE AND SUBTITLE
Modeling and Reconstruction Algorithms for Detection, Location, and Identification of Subsurface Anomalies

5. FUNDING NUMBERS
DAAG29-85-K-0045

6. AUTHOR(S)
Constantine A. Balanis and Mark Frank

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Arizona State Univ
Dept. of Electrical Engineering
Tempe, AZ 85287-7206

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ARO 21470.11-GS

11. SUPPLEMENTARY NOTES
The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
Techniques for modeling and reconstructing images of subsurface anomalies, in a hole-to-hole system configuration, have been investigated. The region under examination is scanned by transmitting electromagnetic rays through the region. Numerical algorithms have been developed for solving the reconstruction problem. The singular value decomposition method has played a key role in fulfilling this goal. In particular, the conjugate gradient algorithm is used with the gradient projection method to obtain a highly stable and efficient technique which is able to incorporate inequality constraints. These inequality constraints represent one type of a priori information that can be used to yield better solutions to detect, locate, and identify subsurface anomalies.
The method of weighted least squares is used to include a priori information into the solution. Two ways of explicity adding this information into the reconstruction process are given, and their effectiveness to reducing additive noise is demonstrated. To obtain images of subsurface anomalies which are the most representative of the actual cross sections being investigated, both continuous wave and time-of-flight measurements are utilized.

14. SUBJECT TERMS

15. NUMBER OF PAGES
274

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

CHAPTER Page

CHAPTER

APPENDIX

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure | Page

Figure		Page

Figure | Page

Figure		Page

Figure		Page

Figure | Page

Figure		Page

Figure | Page

CHAPTER I

INTRODUCTION

In this report we consider the process of obtaining an image of the cross section of an object from measurement of rays which pass through the object. This process is referred to as reconstruction of a cross sectional image of an object from its projections. In this context, a projection is defined to be an integration of some parameter of the object along a line through the object. A ray which passes through the object and attenuates along its path as a function of the object's physical characteristics (along that path) would give an approximation of this line integral.

Practical application of the method of reconstructing a cross section of an object from measurement of a set of line integrals through the object dates back to the 1950s with Bracewell's work in radio astronomy [1]. The application of this theory to medical imaging dates back to the 1960s [2], [3] with the first commercially available scanner introduced in 1972 [4].

Surprisingly, the mathematical foundations for this theory date back to 1917 with Radon's classic paper on recovering a function from its line integrals [5], and to 1923 with Hadamard's work on inverse and ill-posed problems [6]. We call this reconstruction process an inverse problem since we are supplied the projection values and must invert the data in order to obtain some characteristic (as a function of position) of the object being scanned.

Reconstructing images from projections has also found application in nondestructive testing [7] and seismology [8]. However, in this report the emphasis is on scanning an underground cross section of the earth using electromagnetic rays. This operation is often referred to as *geotomography*. One particular measurement geometry which is used for scanning an underground region is shown in Fig. 1.1. In this

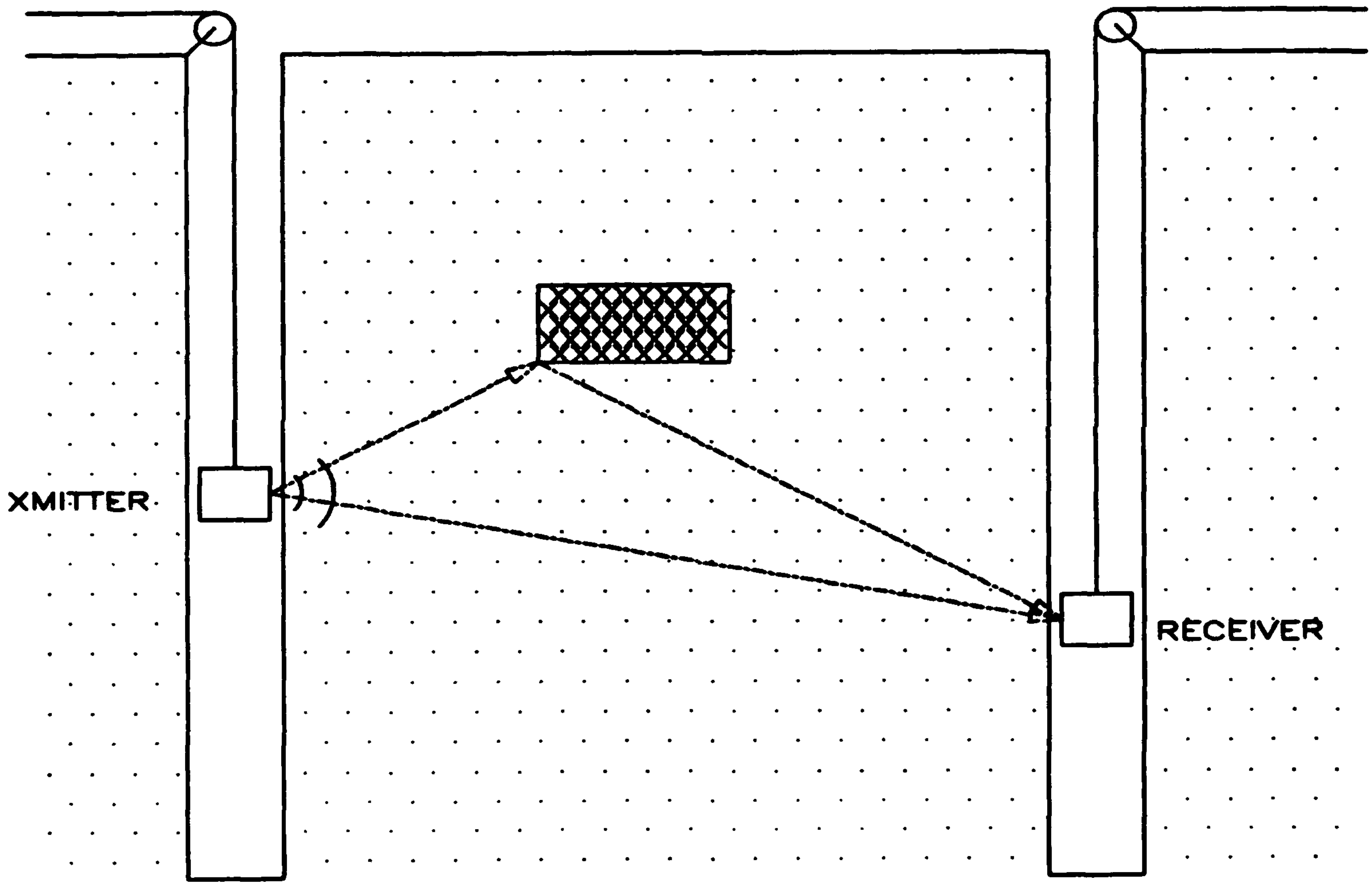


Fig. 1.1. Illustration of obtaining projection data in a cross-borehole geometry.

geometry, two boreholes are drilled into the ground outside of the region to be scanned. A transmitting antenna is lowered into one borehole, and measurements of the received signal are taken in the opposite borehole. These measurements are taken over a range of transmitter and receiver positions. This operation is normally automated using a computer.

The ability to obtain a cross section of a region of earth has important consequences. For example, geotomography has been used to locate burn fronts in coal seams [9], to monitor heavy oil recovery [10], and to detect, locate and identify tunnels [11]. It is apparent that as our natural resources are diminished, better ways of exploring the underground environment will be needed. Geotomography is one method which will help fill this need.

Some of the early work in this field was done at Lawrence Livermore Laboratory [11], [12], [13]. Utilizing a system configuration that measured the power at receiver locations, the attenuation through the earth as a function of position could be found. In obtaining the cross sectional images of attenuation, the assumption was made that the electromagnetic rays follow a direct path from the transmitting to the receiving antenna. This assumption caused errors in the reconstructed images when the rays experienced reflections, refractions, and diffractions. Attempts were made [14] - [18], to reduce the effects of noise, reflections, and refractions by incorporating ray tracing into the geotomography process. This process involves reconstructing an image assuming direct rays, and then iteratively improving the ray path knowledge by using Snell's law to find the path a ray would follow through the current estimate of the image. Unfortunately, this method ignores diffraction of the rays which in some cases may be the dominant effect.

A method was devised by Devaney and others [19], [20] for implicitly incorporating reflections, refractions, and diffractions into the reconstruction process. This method is known as diffraction tomography.

Reconstructions are obtained by inverting the wave equation using either the Born or Rytov approximations. Unfortunately, these approximations are only valid for cross sections containing weak scatterers (slight inhomogeneities), so that this method may not be useful, for example, in tunnel detection. One of the goals of this project was to develop a method which is sufficiently robust to handle cross sections containing strong inhomogeneities, yet still account for diffraction effects. An outline of the development of such a method follows.

Chapter II gives some of the more common models for describing the propagation of electromagnetic waves in the earth. In addition, a new ray optics model [21] is presented which is of great importance to certain reconstruction methods. A thorough understanding of the propagation of electromagnetic waves in earth is a necessary first step for considering the subsurface image reconstruction problem. In fact, the ray optics model developed in this chapter leads to a potentially powerful reconstruction method which is described in Chapter III. This ray optics model along with some of the other models developed in Chapter II serves as building blocks for the reconstruction methods developed here. In addition, some of the difficulties with inverse problems are described in Chapter III with emphasis on image reconstruction theory.

Numerical algorithms for the image reconstruction problem are developed in Chapter IV. A new method for incorporating *a priori* information into the reconstruction process using weighted least squares is presented. An extension of the method of conjugate gradients and an implementation of the singular value decomposition are applied to the subsurface reconstruction problem. It is shown that the conjugate gradient method (for subsurface detection, location and identification) is far superior to the previous standard method used in geotomography, the *algebraic reconstruction technique*, in terms of fast convergence and immunity to noise.

Chapter V presents some methods of post-processing a reconstructed

image to reduce noise, enhance features, and locate, detect, and identify subsurface anomalies [22]. In addition to some standard results, a new detection scheme is discussed. This detection scheme is shown to be most effective in accurately locating high contrast subsurface anomalies, such as tunnels.

A new method of combining reconstructions from two types of measurement processes in order to detect, locate, and identify anomalies in a homogeneous earth is presented in Chapter VI. In this chapter, it is shown that reconstructions obtained using either continuous wave or time-of-flight measurements can lead to ambiguous interpretations (see [23] for a description of time-of-flight reconstructions). A way of avoiding this ambiguity, by utilizing both sets of measurements, is presented along with some guidelines for identifying subsurface anomalies in a cross section of the earth.

CHAPTER II

GEOPHYSICAL MODEL

2.1 Introduction

In order to adequately understand and model the geophysical reconstruction problem, it is first necessary to have a good description of how electromagnetic waves travel through the earth. To this end, the present chapter will attempt to give such a description. The following section will describe the electrical parameters of earth, and their various relationships. The succeeding sections will discuss different ways that the forward problem can be solved. This forward problem consists of generating simulated data of the electromagnetic field at different depths in the receiver borehole given the characteristics of the earth between the receiver and transmitter borehole. This modeling problem is important since:

a) actual field data are available only for limited geological structures,

b) analysis of simulated data gives insight into the inversion process, and

c) analysis of simulated data can help to optimize the field measurement process.

After the forward problem is well understood, the reconstruction problem can then be solved.

2.2 Electromagnetic Characteristics of the Earth

Unless otherwise noted, this section will consider the problem of determining the electromagnetic fields in a homogeneous earth. This means that the electric parameters do not change with location in the earth. This simplified assumption will be relaxed in the next section.

2.2.1 Electrical Parameters for Continuous Wave Tomography

The basic electrical parameters of interest in describing a material are:

a) ε - permittivity, in Farads/meter,

b) μ - permeability, in Henries/meter, and

c) σ - conductivity, in Siemens/meter.

It is customary to express the permittivity and permeability in terms of their values in a vacuum as

$$\varepsilon = \varepsilon_r \, \varepsilon_0, \tag{2-1}$$

$$\mu = \mu_r \, \mu_0, \tag{2-2}$$

where ε_0 (=8.854×10^{-12}F/m) and μ_0 (=$4\pi\times10^{-7}$H/m) are the free space values. For the cases we will be considering, it will be assumed that the earth does not contain any magnetic materials (e.g. iron ore deposits); therefore $\mu_r = 1$. On the other hand, the relative permittivity, ε_r, can take on values much greater than 1, depending on such conditions in the ground as type of rock/soil, particle sizes, and water content. The conductivity of the earth also depends on these conditions.

Although the relative permittivity, ε_r, and conductivity, σ, will, in general, depend on frequency [24],[25], for the following we will assume that these two parameters are independent of frequency (for illustrative purposes). Given this assumption, we see that σ, ε, and μ are static parameters. Since we are interested in the propagation of waves in the ground, a more useful quantity is the propagation constant

$$\gamma = \alpha + j\beta = [(\sigma + j\omega\varepsilon)j\omega\mu]^{1/2} \tag{2-3}$$

where $j = \sqrt{-1}$, ω is the radian frequency, α is the attenuation constant, and β is the phase constant. This propagation constant is used to describe the transmission of a uniform plane wave of frequency, ω, through a homogeneous medium (Note: an $e^{+j\omega t}$ time convention is assumed). The traveling nature of the wave is evident if we consider a wave moving in the +z direction, then the field at any point along the z axis is given by

$$u(z) = u_0 e^{-\gamma z} = u\ e^{-(\alpha + j\beta)z} \qquad (2\text{-}4)$$

where u_0 is the magnitude of the field at z=0. It is now evident that the field is attenuated in the +z direction according to the constant α.

This attenuation constant is important for cross-borehole tomography since it determines the amount of electromagnetic power which reaches the receiver borehole. Note that α is not truly a constant in that it will depend on ω, σ, ε, and μ as

$$\alpha = \omega\sqrt{\mu\varepsilon}\left\{\frac{1}{2}\left[\ 1 + \left(\frac{\sigma}{\omega\varepsilon}\right)^2\ \right]^{1/2} - 1\right\}^{1/2}, \qquad (2\text{-}5)$$

and σ and ε will, in general, be functions of position.

It is useful to get a good characterization of attenuation, since, when continuous wave (CW) geotomography is performed, the attenuation of the earth as a function of position is the unknown quantity to be found. This will allow us to map a cross section of the earth using attenuation as our parameter. Also, the attenuation will determine how far the electromagnetic waves can travel in the earth and still be measured at the receiver. For these reasons, in Figs. 2.1 - 2.3 the attenuation is plotted as a function of frequency, conductivity, and relative permittivity. In generating these plots, all other parameters

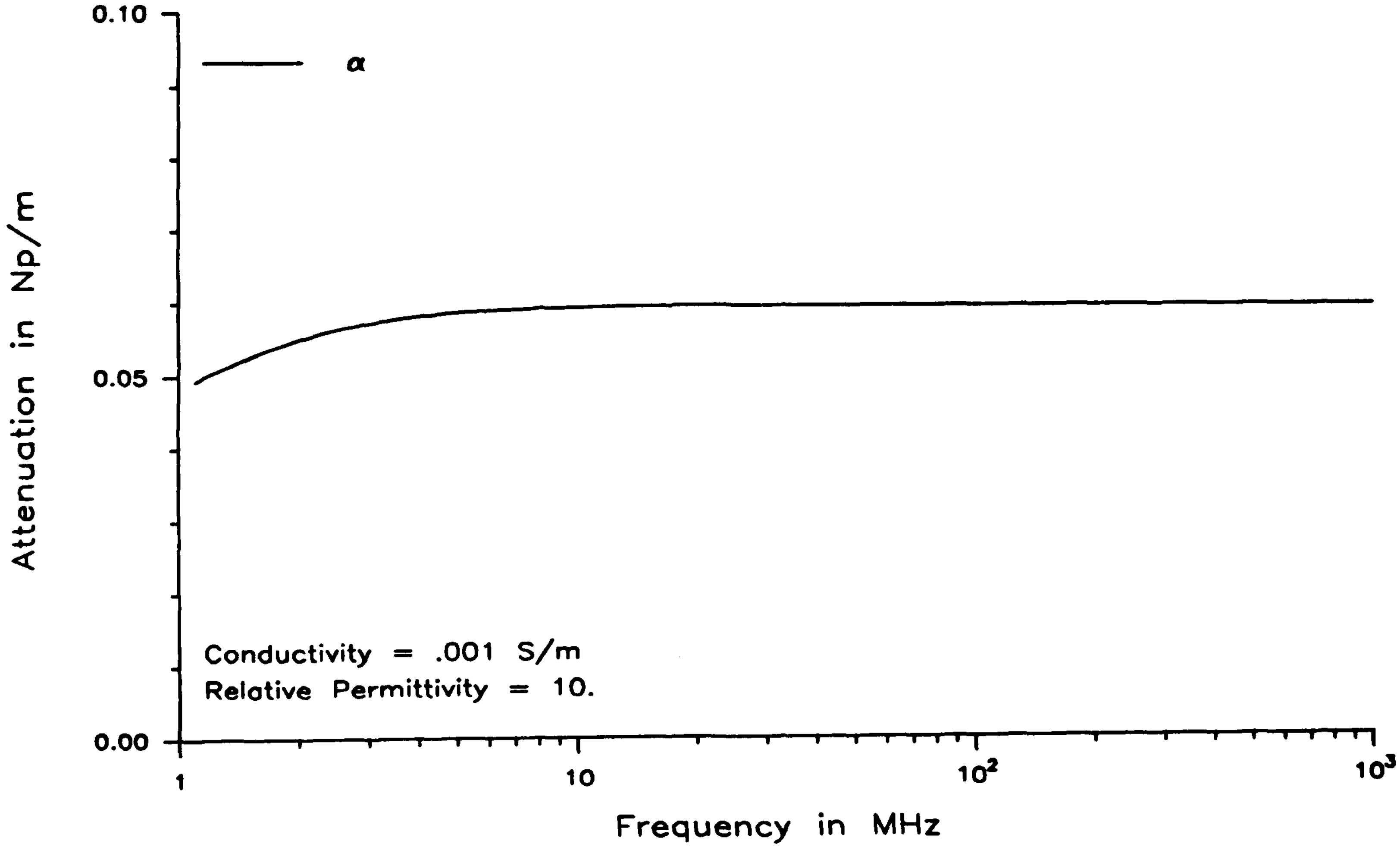


Fig. 2.1. Attenuation constant versus frequency for permittivity = 10, conductivity = 0.001 S/m.

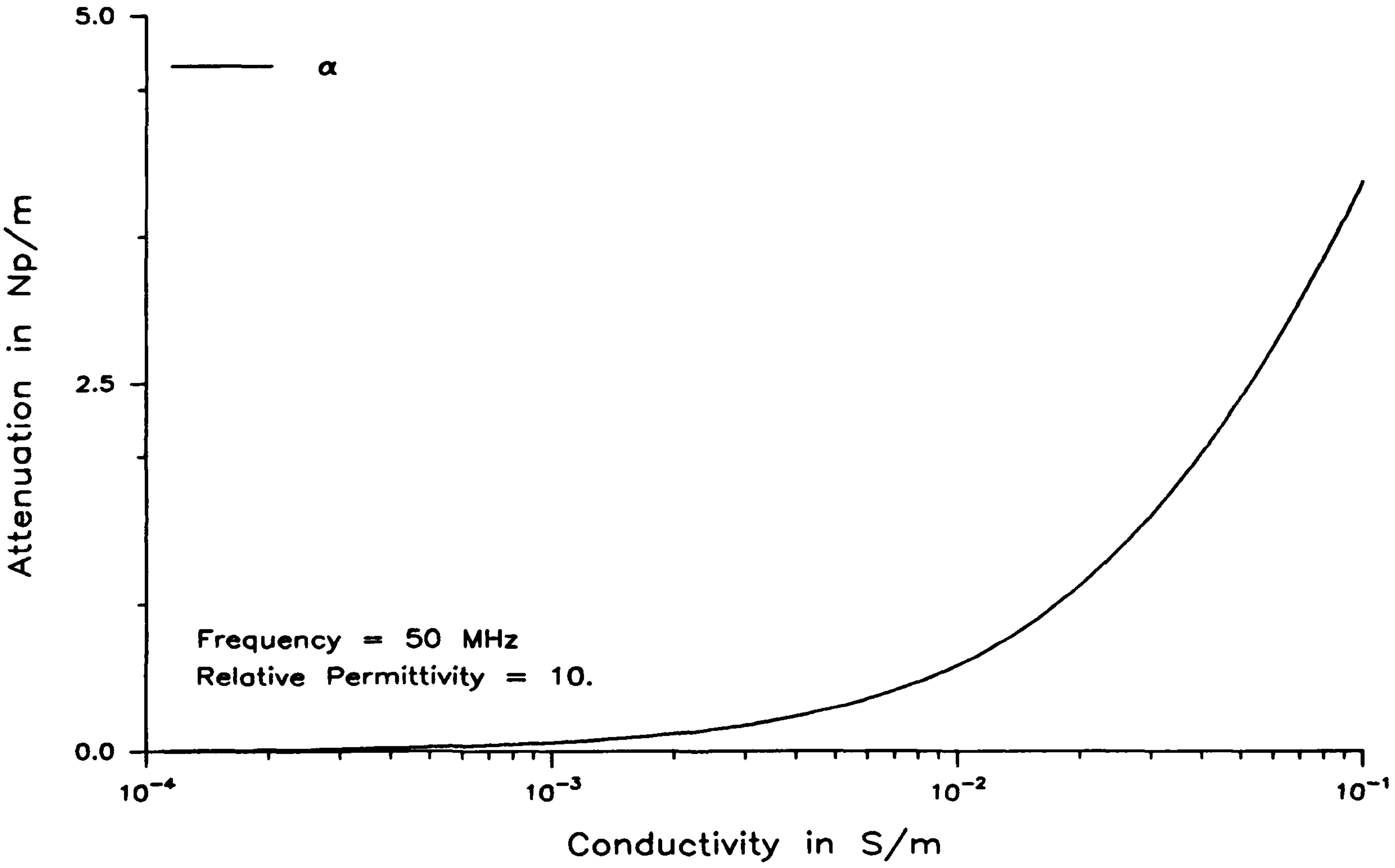


Fig. 2.2. Attenuation constant versus conductivity for permittivity = 10, frequency = 50 MHz.

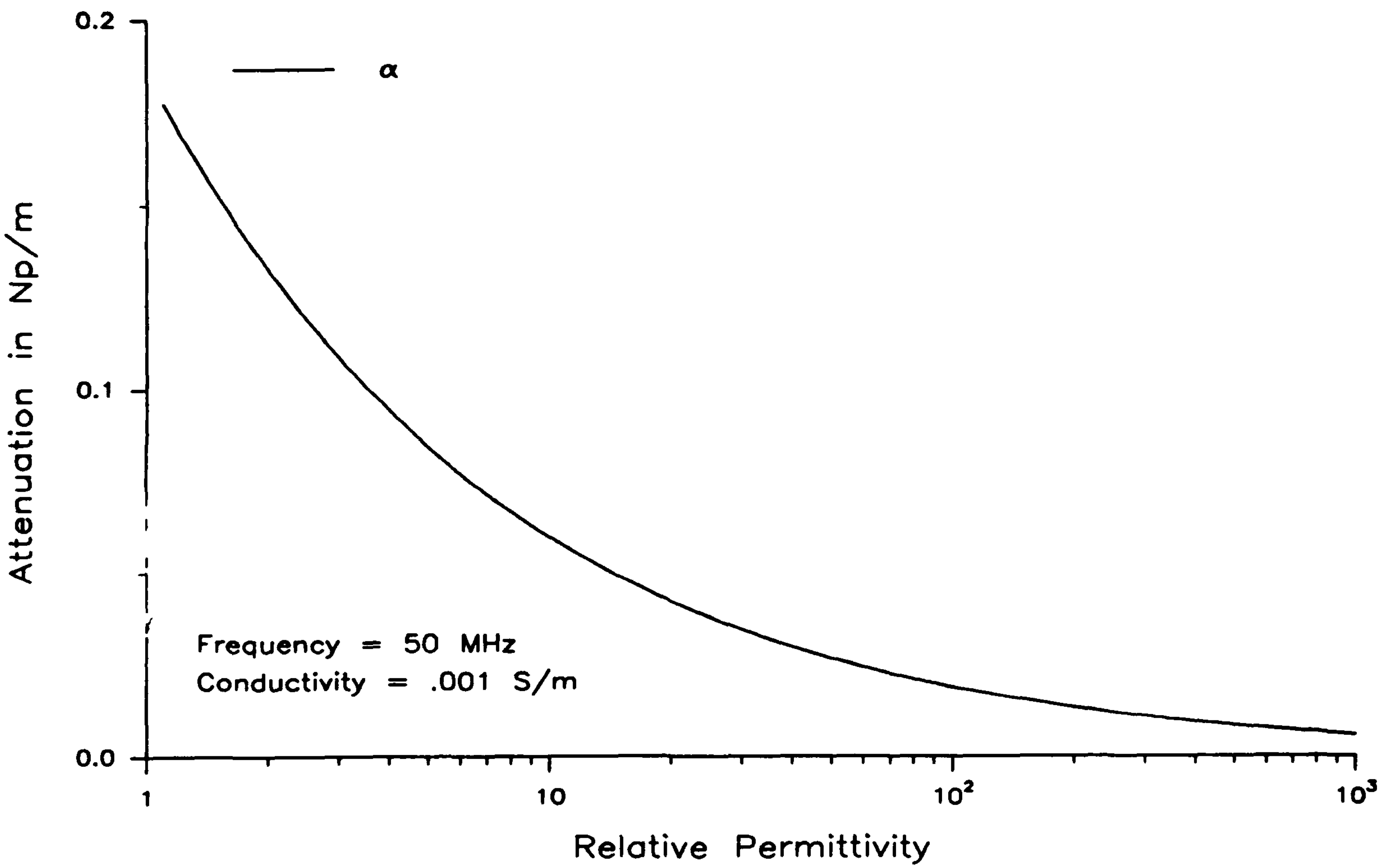


Fig. 2.3. Attenuation constant versus permittivity for conductivity = 0.001 S/m, frequency = 50 MHz.

were kept constant at values representative of dry soil (ε_r = 10, σ = 0.001 S/m). The following conclusions may be drawn from examination of these figures:

a) Fig. 2.2 gives an upper limit on the conductivity of the earth which can be scanned using a transmission frequency of 50 MHz. For example, for σ = 0.2 S/m the attenuation is about 1 Np/m. If we assume a transmitted power of 1000 W, then at a distance of 10 meters from the transmitter the received power would be on the order of 2 μW, which would be difficult to measure.

b) Finally, Fig. 2.3 shows that attenuation is also a strong function of permittivity.

These conclusions suggest that the region of earth to be scanned for geotomography purposes needs to be investigated before any measurements are made.

Another parameter of interest for a traveling wave is wavelength, given by

$$\lambda = \left\{\frac{\omega^2\mu\varepsilon}{8\pi^2}\left[\ 1 + \left(\frac{\sigma}{\omega\varepsilon}\right)^2\ \right]^{1/2} + 1\right\}^{-1/2} \text{m}. \tag{2-6}$$

The wavelength gives some indication of the resolution of the geotomography process. For example, we do not expect to "see" objects which are significantly smaller than the intrinsic wavelength. In Figs. 2.4 - 2.6, the wavelength in meters is plotted versus the same quantities as the attenuation was in Figs. 2.1 - 2.3. Fig. 2.4 shows that measurements must be made at frequencies greater than 10 MHz, for the given values of conductivity and permittivity, in order to resolve objects smaller than 10 m. This suggests a trade-off between increasing the transmission frequency to increase resolution (Fig. 2.4), and decreasing frequency to obtain greater penetrating range (Fig. 2.1).

Fig. 2.5 shows the wavelength as a function of conductivity with

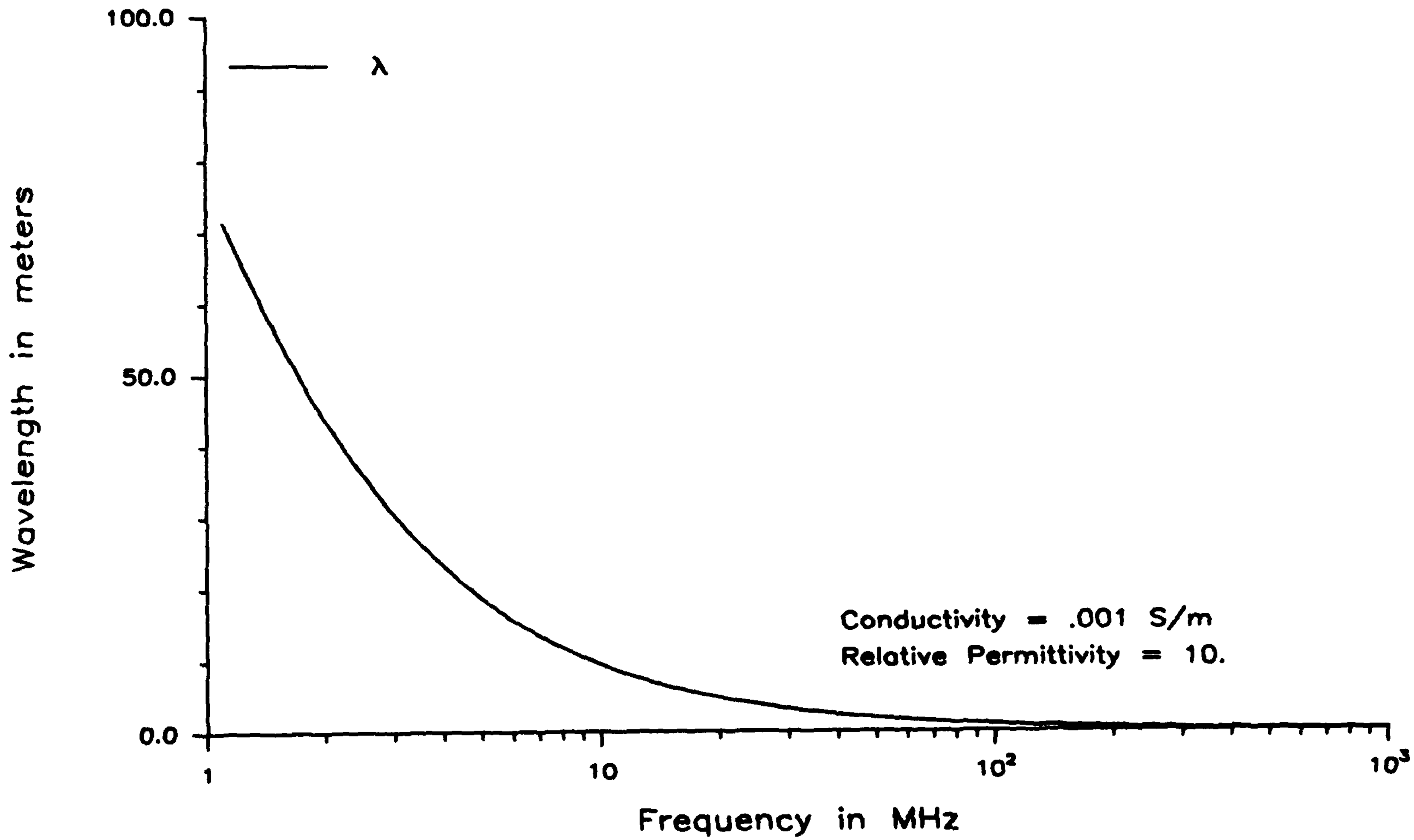


Fig. 2.4. Wavelength versus frequency for permittivity = 10, conductivity = 0.001 S/m.

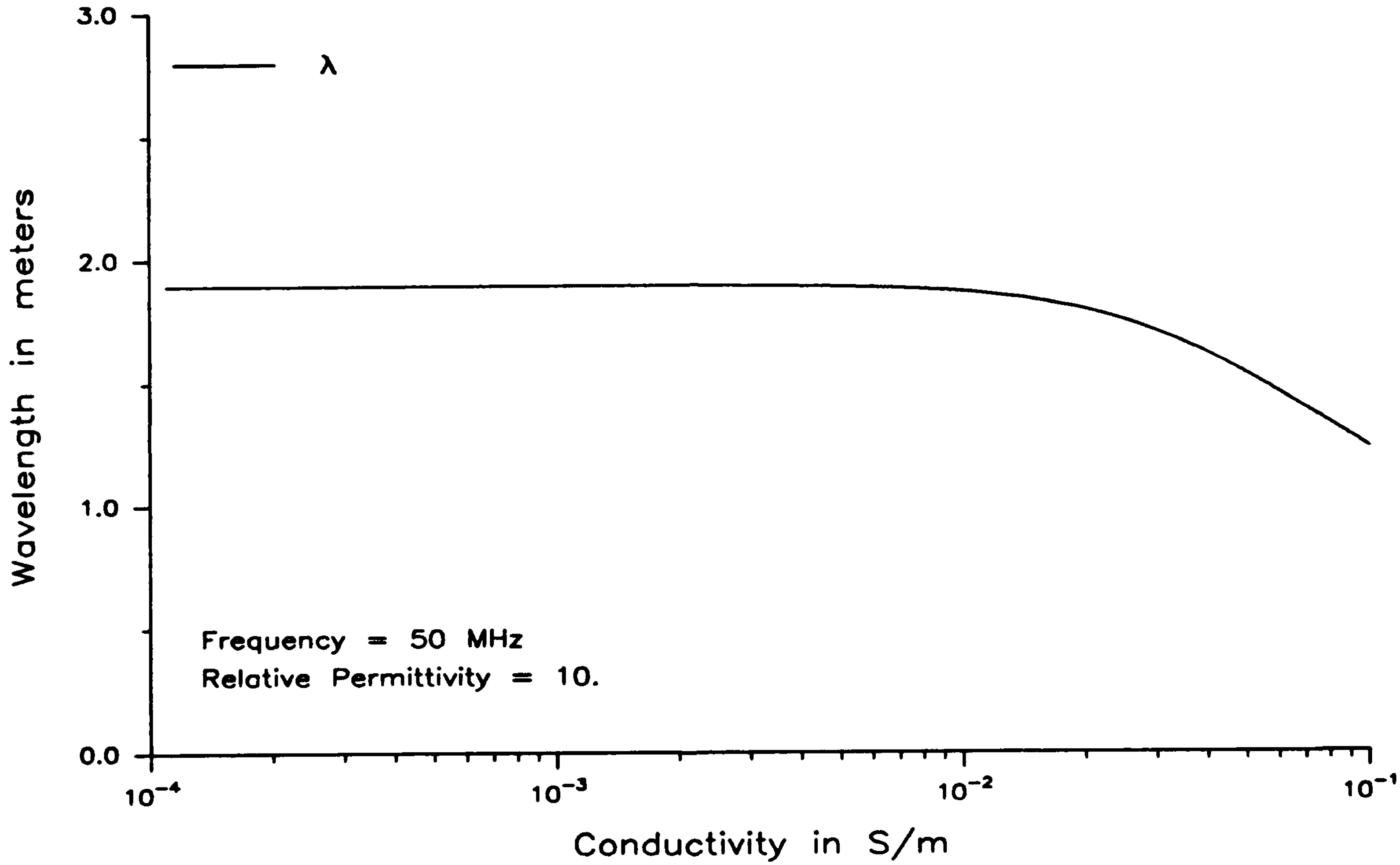


Fig. 2.5. Wavelength versus conductivity for permittivity = 10, frequency = 50 MHz.

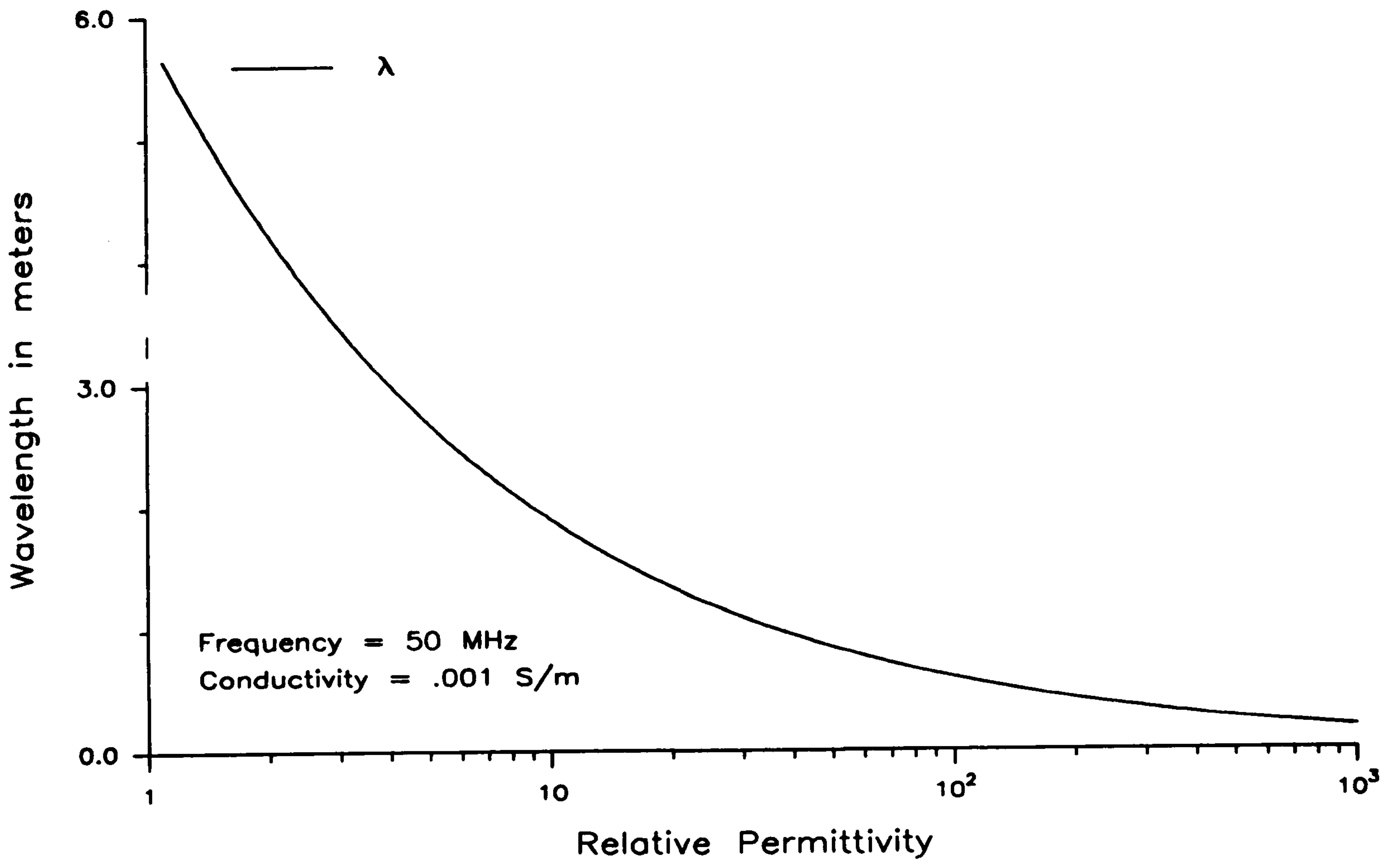


Fig. 2.6. Wavelength versus permittivity for conductivity = 0.001 S/m, frequency = 50 MHz.

frequency and permittivity maintained constant. Finally, Fig. 2.6 is a plot of wavelength versus permittivity with frequency maintained constant. These figures will be referred to later, when sample reconstructions are presented.

2.2.2 Electrical Parameters for Time-of-Flight Measurements

The main electrical parameter of interest for time-of-flight (TOF) measurements is the wave velocity, given by

$$\nu = \left\{\frac{\mu\varepsilon}{2}\left[\ 1 + \left(\frac{\sigma}{\omega\varepsilon}\right)^2\ \right]^{1/2} + 1\right\}^{-1/2} \mathrm{m}. \qquad (2\text{-}7)$$

When taking TOF measurements, the time it takes a pulse to travel from transmitter to receiver is measured. Therefore, the time taken will be a direct function of the intrinsic velocity of the intervening medium. Like the attenuation, the velocity is a function of frequency, conductivity, and relative permittivity. These relationships are plotted in Figs. 2.7 - 2.9. Also plotted in these figures are the velocities that would be obtained if the conductivity were zero. This zero conductivity velocity is important since σ will be neglected when deriving the reconstruction algorithm for TOF measurements. By neglecting the conductivity, it can be seen that the velocity is a function of only the permeability and permittivity. If the permeability is assumed to be constant, the TOF measurements will then allow us to map the permittivity of a cross section of the earth versus position. As can be seen in Fig. 2.7, σ can be neglected when the conductivity is equal to 0.001 S/m and the frequency is greater than 10 MHz. Figs. 2.8 and 2.9 also give justification for using this approximation.

2.3 Sinusoidal Response of a Homogeneous Earth Containing Isolated

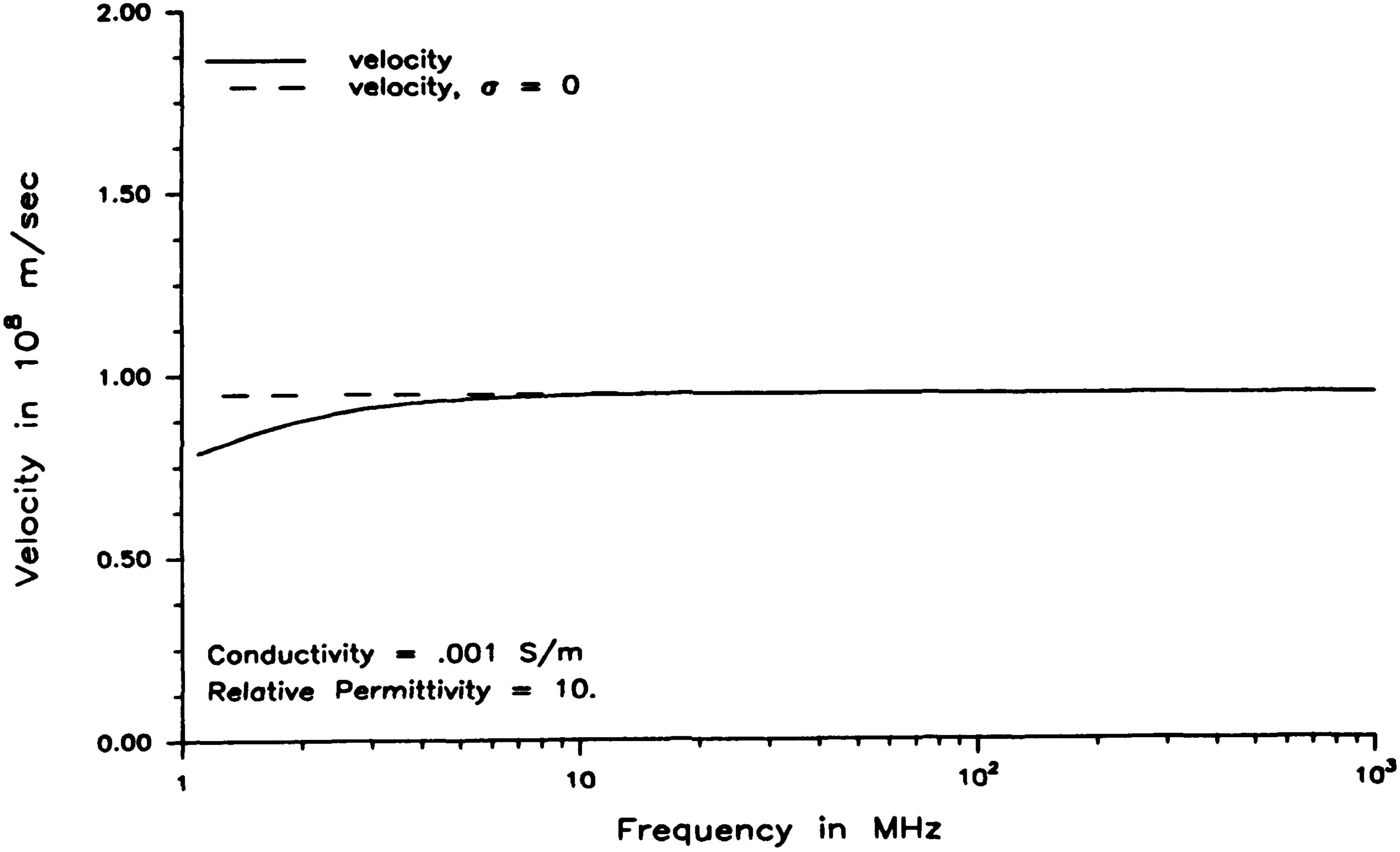


Fig. 2.7. Wave velocity versus frequency for permittivity = 10, conductivity = 0.001 S/m.

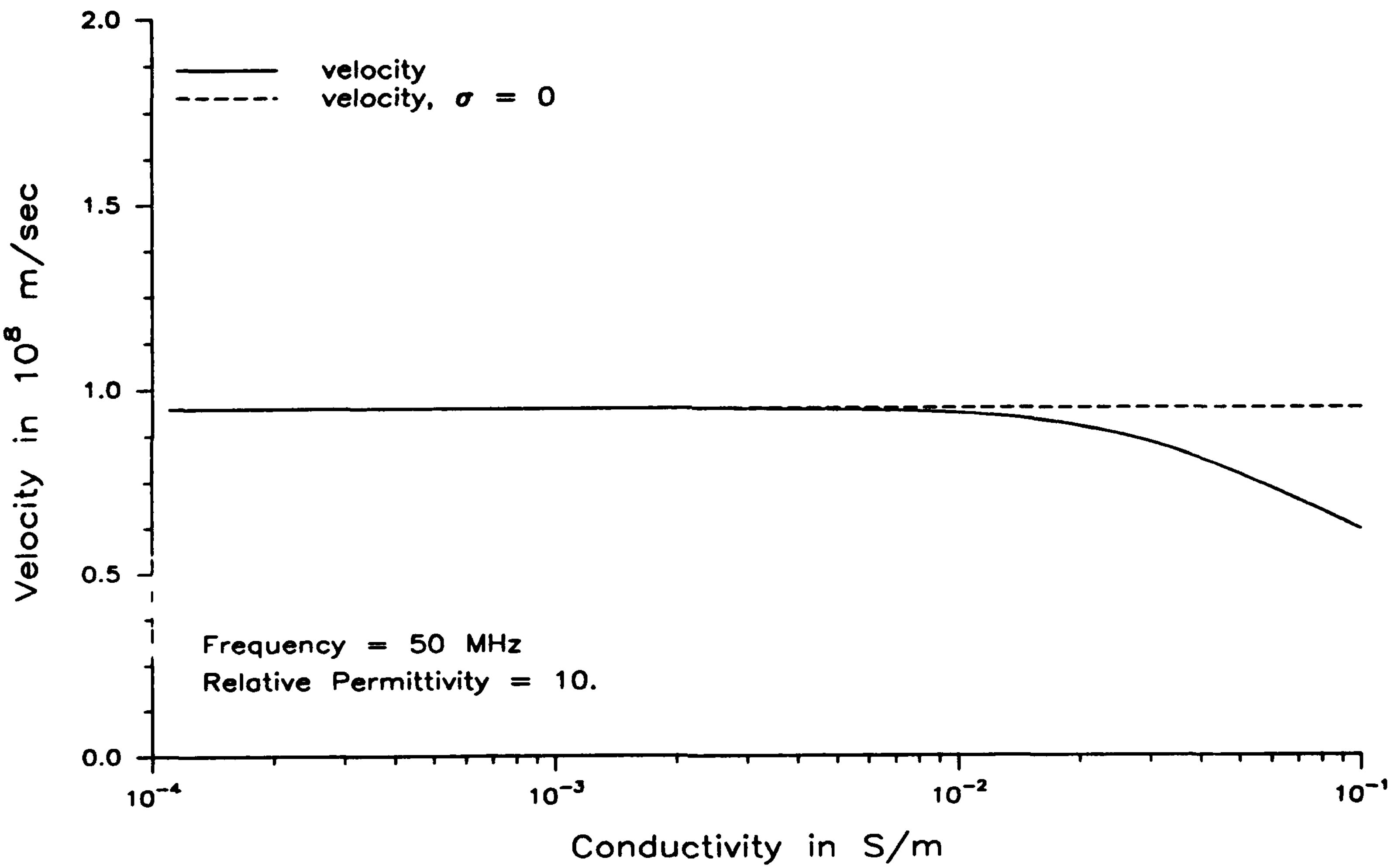


Fig. 2.8. Wave velocity versus conductivity for permittivity = 10, frequency = 50 MHz.

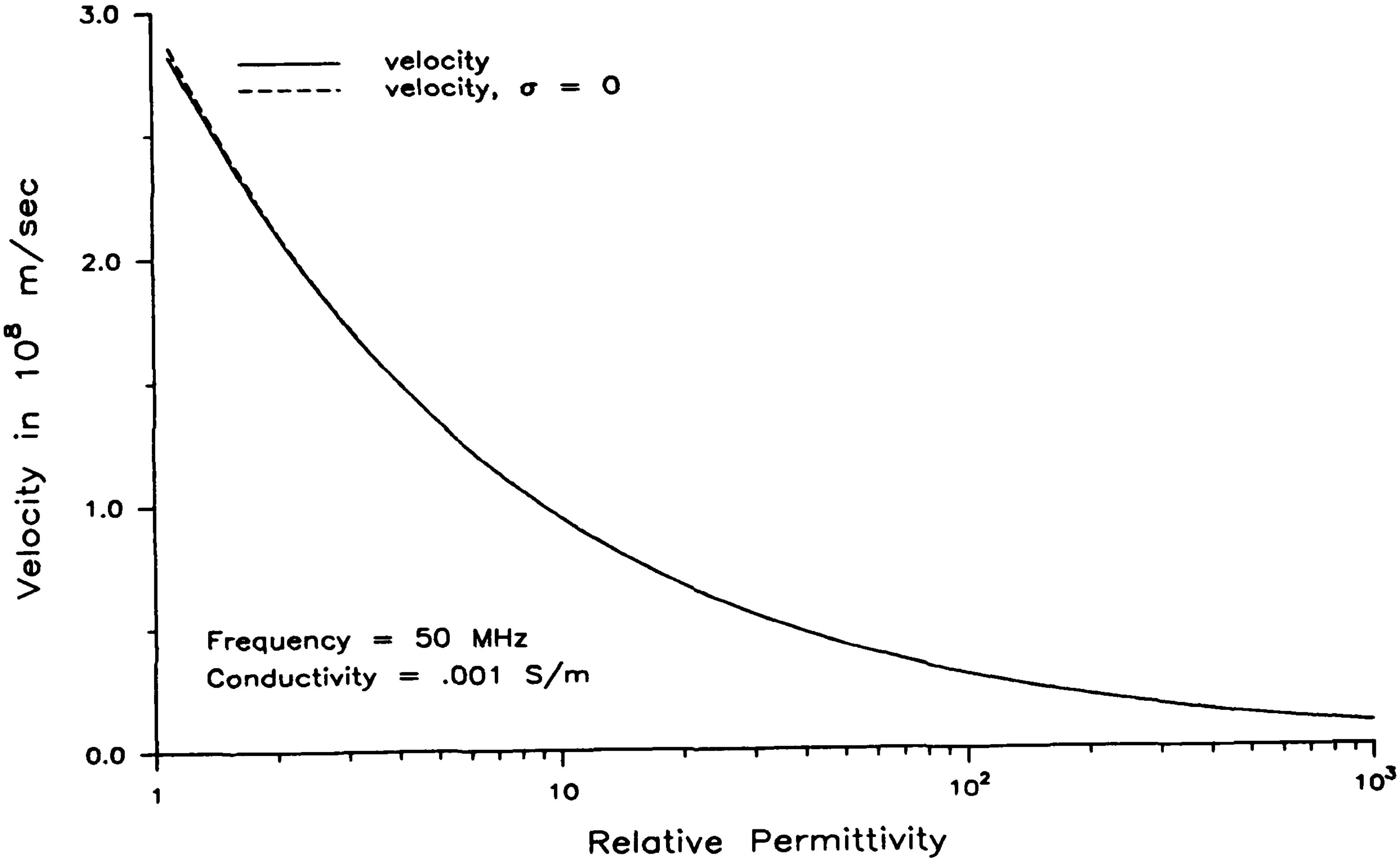


Fig. 2.9. Wave velocity versus permittivity for conductivity = 0.001 S/m, frequency = 50 MHz.

Anomalies

2.3.1 Introduction

Up until this point in the chapter, the discussion has been for the case of an earth whose electrical parameters do not change with position. Of course, if this were always the case, there would be no need for the theory of geotomography.

Heterogeneities can exist in the earth in the form of: rock layers, faults, seams, extrusions, water deposits, or tunnels. We will refer to isolated heterogeneities of limited extent such as water deposits or tunnels as anomalies. These anomalies will be easier to detect than the other types of heterogeneities because of their finite extent in the viewing space. In fact, it is often the case that the geophysicist attempts to detect and locate anomalies in spite of the presence of other heterogeneities. So, although theory exists for modeling electromagnetic waves in stratified media [26], it is the opinion of this author, that these phenomena must be handled on a case-by-case basis. For instance, if a geophysicist is exploring for oil, and s/he knows that a layer of limestone exists in the region to be examined, s/he should then adapt the model accordingly.

Because these global heterogeneities should be addressed only when there is *a priori* knowledge about their presence, and because they will in general complicate the geophysical model, they will not be discussed in detail here. Rather, in the remainder of this chapter, modeling of isolated anomalies in a homogeneous earth will be presented.

We consider in this section the sinusoidal response of the earth containing a line source antenna. This theory will be important for considering electromagnetic probing of the earth for continuous wave (power) measurements. In the next section the time response of the earth will be investigated in order to characterize the time-of-flight measurement process.

2.3.2 Green's Function Solution for a Homogeneous Earth

It is important to derive a Green's function solution for a homogeneous earth. This Green's function gives the electromagnetic response of an antenna radiating in a homogeneous earth. Once this solution is found, it can be used to solve more complex problems in which the earth contains scatterers. These more complex problems will be given attention later in this section. Fortunately, this Green's function solution exists [27], [28], for the so-called damped wave equation given by

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} - \mu\sigma\frac{\partial E}{\partial t} - \mu\varepsilon\frac{\partial^2 E}{\partial t^2} = \mu\frac{\partial J}{\partial t} , \qquad (2\text{-}8)$$

where E is the electric field assumed to be a function of x, y, and time; and J is the current density (J will be non-zero only at the location of the electric line source which will be aligned along the z direction). This equation relates the electric field in a two dimensional region to the position in a region, the time, and the derivative of the source. Fig. 2.10 is an illustration of the problem. Equation (2-8) is a result of combining the equations

$$\nabla \times E = -\frac{\partial}{\partial t}\mu H , \qquad (2\text{-}9a)$$

$$\nabla \times H = J + \sigma E + \frac{\partial}{\partial t}\varepsilon E , \qquad (2\text{-}9b)$$

$$\nabla \circ E = 0 \text{ , and} \qquad (2\text{-}9c)$$

$$\nabla \times \nabla \times E = \nabla(\nabla \circ E) - \nabla^2 E ; \qquad (2\text{-}9d)$$

in the following manner.

a) Take the curl of (2-9a).

b) Use (2-9c) and (2-9d) in the resulting equation.

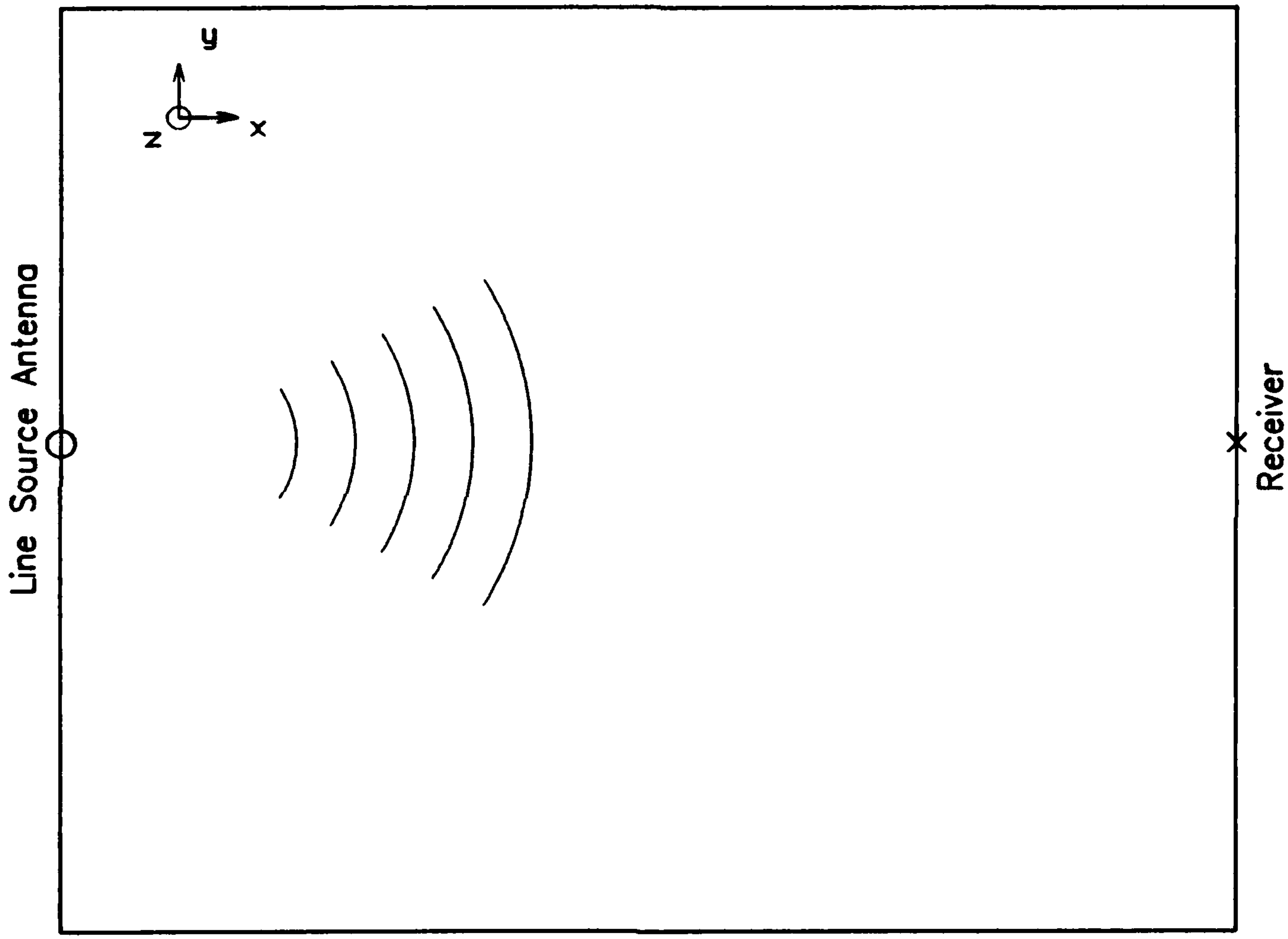


Fig. 2.10. Configuration for the time response of the electric field using a line source antenna.

c) Use (2-9b) to substitute for $\nabla \times H$ in the result.

Equations (2-9 a, b, and c) are three of Maxwell's equations, H is the magnetic field intensity, and J is the electric current density representing the source.

The Green's function can be found by proceeding as follows [27],

a) Replace the right hand side (the source term) of the damped wave equation by the three dimensional dirac delta function, $-\delta(x)\delta(y)\delta(t)$. In this way it is assumed that the source is at the origin of the coordinate system.

b) Apply a two-dimensional spatial Fourier transform to the resulting equation, to get

$$-k^2 U - \mu\sigma\frac{\partial U}{\partial t} - \mu\epsilon\frac{\partial^2 U}{\partial t^2} = \frac{\delta(t)}{2\pi}, \qquad (2\text{-}10)$$

where we have defined the quantity U to be

$$U = U(k_1, k_2, t) := \mathbf{F}_{xy} G(x,y,t)$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(x,y,t)\, e^{-jk_1 x} e^{-jk_2 y}\, dx\, dy \qquad (2\text{-}11)$$

In this equation, $k^2 := k_1^2 + k_2^2$, k_1 and k_2 are spatial frequency variables, **F** is the Fourier transform operator, and G has been substituted for E to denote that the solution is a Green's function. Implicit in this development is that the solution is of such form (for example, E and its first two derivatives are square integrable) to allow taking such liberties as bringing the time derivatives outside of the Fourier integral.

c) As can be seen, (2-10) is an ordinary differential equation, whose solution is again a Green's function (note the delta function on the right hand side). This equation can be solved by again using a Fourier transform, but with respect to the

time variable to get

$$-k^2V - j\mu\sigma\omega V + \mu\varepsilon\omega^2 V = (2\pi)^{-3/2} \qquad (2\text{-}12)$$

where $V = V(k_1,k_2,\omega) := \mathbf{F}_t U(k_1,k_2,t)$ and ω is the temporal frequency. This equation can be solved algebraically to yield

$$V = \frac{(2\pi)^{-3/2}}{k^2 + j\mu\sigma\omega - \mu\varepsilon\omega^2} = \frac{-(2\pi)^{-3/2}}{(\omega - \omega_1)(\omega - \omega_2)} \qquad (2\text{-}13a)$$

where ω_1 and ω_2 are the roots of the denominator given by

$$\omega_1 = j\Gamma + p \qquad (2\text{-}13b)$$

$$\omega_2 = j\Gamma - p \qquad (2\text{-}13c)$$

$$\Gamma = \frac{\sigma}{2\varepsilon}\ , \ p = \left[\frac{k^2}{\mu\varepsilon} - \Gamma^2\right]^{1/2}. \qquad (2\text{-}13d)$$

Note that these roots are in the upper half of the complex plane.

d) To find $U(k_1,k_2,t)$ from (2-12), take the inverse Fourier transform as

$$\begin{aligned} U &= \mathbf{F}_t^{-1} V \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{-(2\pi)^{-3/2}}{(\omega - \omega_1)(\omega - \omega_2)} e^{j\omega t} d\omega\ . \end{aligned} \qquad (2\text{-}14)$$

This integral can be solved by using the residue theorem and Jordan's lemma to obtain

$$U(k_1,k_2,t) = \frac{e^{-\Gamma t}}{2\pi p}\sin(pt), \text{ for } t>0, \qquad (2\text{-}15)$$

where Γ and p are defined in (2-13).

e) To find $G(x,y,t)$ from (2-15), again inverse transform as

$$G = F_{xy}^{-1}U$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{e^{-\Gamma t}}{2\pi p}\sin(pt)\; e^{jk_1x}e^{jk_2y}\; dk_1dk_2 \qquad (2\text{-}16)$$

f) It is convenient now to convert to polar coordinates such that

$$\rho^2 = x^2 + y^2 , \qquad (2\text{-}17a)$$

$$k^2 = k_1^2 + k_2^2 , \qquad (2\text{-}17b)$$

$$e^{jk_1x}\, e^{jk_2y} = e^{jk\rho\cos\Phi} \qquad (2\text{-}17c)$$

$$dk_1dk_2 = k\; dk\; d\varphi \qquad (2\text{-}17d)$$

where $\cos\theta = x/\rho$, and $\Phi = \varphi - \theta$. Equation (2-16) becomes

$$G = F_{\rho\varphi}^{-1}U$$

$$= \frac{1}{2\pi}\int_0^{\infty}\int_0^{2\pi}\frac{e^{-\Gamma t}}{2\pi p}\sin(pt)\; e^{jk\rho\cos\Phi}\, k\; d\varphi\; dk$$

$$= \frac{1}{2\pi}\int_0^{\infty}\frac{e^{-\Gamma t}}{p}\sin(pt)\; J_0(k\rho)\; k\; dk$$

$$= \frac{\mu\varepsilon e^{-\Gamma t}}{2\pi}\int_0^\infty \frac{\sin t(k^2-\Gamma^2)^{1/2}}{(k^2-\Gamma^2)^{1/2}}\, k\, J_0(k\rho\sqrt{\mu\varepsilon})\, dk, \qquad (2\text{-}18)$$

where $J_0(\circ)$ is the Bessel function of zero order. This integral can be evaluated by referring to the table of Hankel transforms in [29] to get

$$G(\rho,t) = \frac{\mu\varepsilon e^{-\Gamma t}}{2\pi}\, \frac{\cosh[\Gamma(t^2-\rho^2\mu\varepsilon)^{1/2}]}{(t^2-\rho^2\mu\varepsilon)^{1/2}}\, H(t-\rho\sqrt{\mu\varepsilon}), \qquad (2\text{-}19)$$

where $H(\circ)$ is the Heaviside step function, which is equal to zero for its argument less than zero, and is unity for its equal or greater than zero.

The result derived above is the electric field at a radial distance, ρ, and time, t, assuming a source term of

$$J = \delta(x)\delta(y)[1 - H(t)], \qquad (2\text{-}20a)$$

$$\frac{d}{dt}[1 - H(t)] = -\delta(t), \qquad (2\text{-}20b)$$

where the derivative is taken in the distributional sense. Now that this basic problem is solved, we can determine the sinusoidal response of a homogeneous earth, and earth containing scatterers.

2.3.3 Sinusoidal Response of Homogeneous Earth

We consider again the damped wave equation developed in the last section. The source term as before is an infinite current-carrying cable in the z direction (refer again to Fig. 2.10). The fields are

invariant in the z direction, so the problem reduces to a two-dimensional situation in (x,y) or (ρ,φ), for cylindrical coordinates. The problem is further simplified if we assume that the current in the cable has a sinusoidal variation of ω rad/s. If the location of the cable coincides with the origin of the coordinate system, then this source term takes the form

$$J = I\ e^{j\omega t}\delta(x)\delta(y), \tag{2-21}$$

where I is the magnitude of the current through the cable. For this sinusoidal source term, the solution will also be sinusoidal of the form

$$E(x,y,t) = E(x,y)\ e^{j\omega t}. \tag{2-22}$$

Now substitute (2-21) and (2-22) into the damped wave equation, take the time derivatives, and then cancel out the $e^{j\omega t}$ to get

$$\nabla^2 E - \gamma^2 E = j\omega\mu I\delta(x)\delta(y). \tag{2-23}$$

Fourier transforming with respect to x and y, and then collecting terms gives

$$F_{xy}E = \frac{-j\omega\mu I}{2\pi(k^2+\gamma^2)}, \tag{2-24}$$

where

$$k^2 = k_1^2 + k_1^2. \tag{2-25}$$

The electric field can now be found in polar coordinates by an inverse transform as

$$E(\rho,\varphi) = \frac{1}{2\pi}\int_0^\infty \int_0^{2\pi} \frac{-j\omega\mu I}{2\pi(k^2+\gamma^2)} e^{jk\rho\cos\varphi}\, k\, d\varphi\, dk$$

$$= \frac{-j\omega\mu I}{2\pi}\int_0^\infty \frac{J_0(k\rho)}{(k^2+\gamma^2)}\, k\, dk$$

$$= \frac{-\omega\mu I}{4} H_0^{(2)}(-j\gamma\rho), \tag{2-26}$$

where the last integral was evaluated by referring to the Hankel transform table in [29]. As would be expected from the symmetry of the problem, the electric field is invariant with respect to φ. Since the electric field is a function only of the distance from the antenna, it can be given in rectangular coordinates as

$$E(\mathbf{X}) = \frac{-\omega\mu I}{4} H_0^{(2)}(-j\gamma|\mathbf{X}|), \tag{2-27}$$

where we have defined

$$\mathbf{X} := (x\ y)^T \tag{2-28}$$

as a two-dimensional vector, and

$$|\mathbf{X}| = (x^2 + y^2)^{1/2}. \tag{2-29}$$

If the cable axis is at $\mathbf{X}_t$, which does not coincide with the origin, the electric field will be

$$E(\mathbf{X}) = \frac{-\omega\mu I}{4} H_0^{(2)}(-j\gamma|\mathbf{X}-\mathbf{X}_t|). \tag{2-30}$$

This completes the discussion of the response of a homogeneous earth containing a line source antenna with a sinusoidal current variation.

2.3.4 Sinusoidal Response of a Homogeneous Earth Containing a Single Circular Cylindrical Anomaly

Refer to the geometry shown in Fig. 2.11. We consider finding the sinusoidal response of a homogeneous earth containing a circular cylinder. A problem of this type is solved [30], [31] by considering the field external to the cylinder to be the sum of an incident field and a scattered field as

$$E = E^i + E^s \ . \tag{2-31}$$

The incident field is the field from the line source antenna assuming the cylinder is not present; it is given by (2-30). The scattered field (E^s) is now expanded in a Fourier series, where the Fourier functions are the eigenfunctions of the homogeneous (no source term) damped wave equation with origin at the center of the cylinder. They are given by

$$\Phi_n(\rho,\varphi) = e^{jn\varphi} J_n(-j\gamma_a\rho), \text{ for } \rho \leq a, \tag{2-32a}$$

$$\Psi_n(\rho,\varphi) = e^{jn\varphi} H_n^{(2)}(-j\gamma_e\rho), \text{ for } \rho \geq a, \tag{2-32b}$$

where a is the cylinder radius. The origin is taken at the cylinder center, and (ρ,φ) are the cylindrical coordinates. The subscripts on the propagation constants are 'a' for anomaly (for inside the cylinder) and 'e' for external (for outside the cylinder). The Bessel functions were chosen for $\rho\leq a$ since they remain finite at $\rho=0$ (and represent standing waves), while the Hankel functions were chosen for $\rho\geq a$ since they decay rapidly as $\rho\to\infty$ (and therfore represent traveling waves). For the conditions under which a function can be expanded in terms of these functions, the reader is referred to [32]. Using the expansion functions given above, the scattered field takes the forms

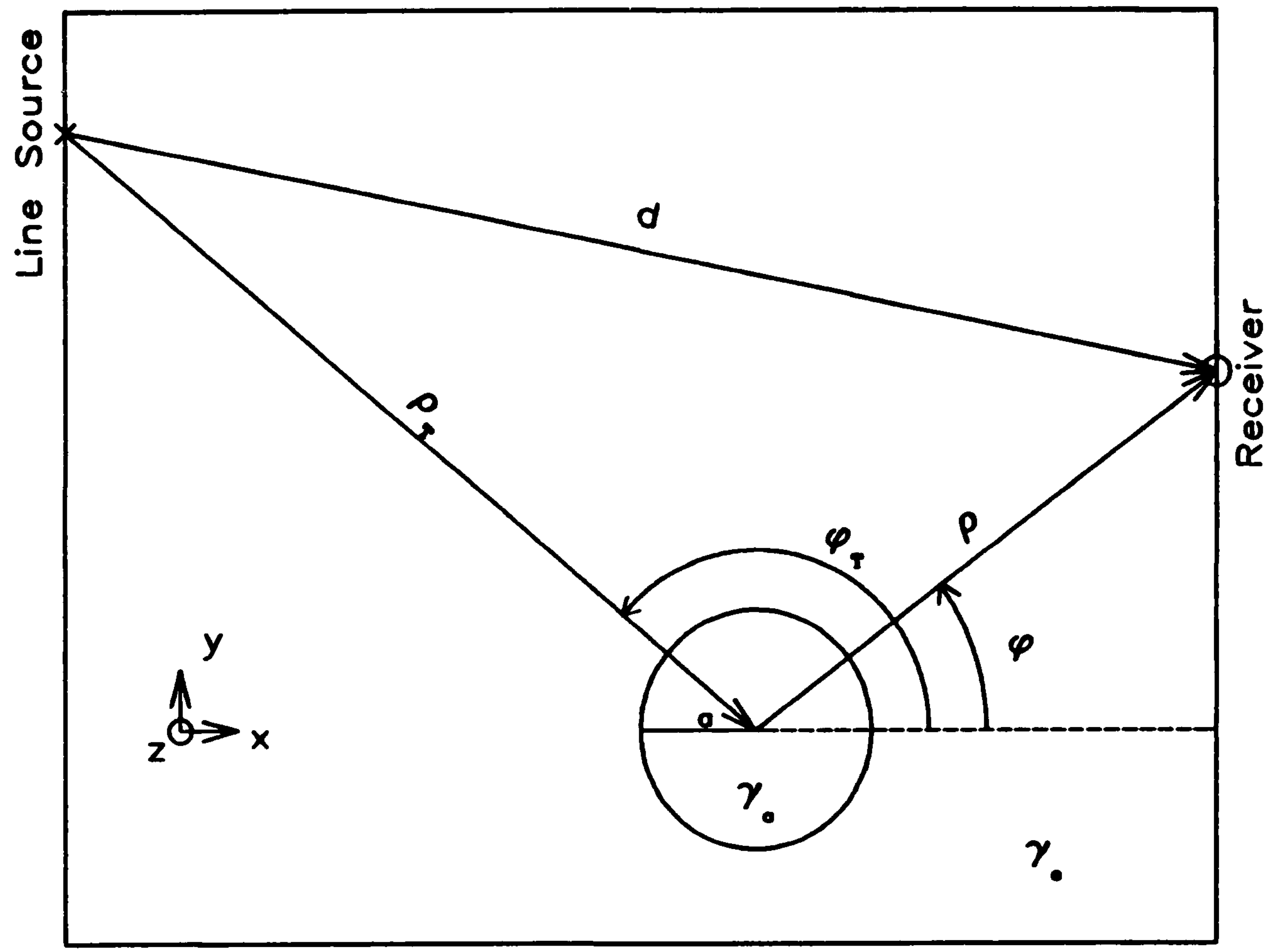


Fig. 2.11. Geometry for the evaluation of the electromagnetic field of a line source antenna in the presence of a circular cylinder.

$$E^s(\rho \leq a) = \sum_{n=-\infty}^{\infty} b_n \Phi_n(\rho, \varphi) \ , \qquad \text{(2-33a)}$$

$$E^s(\rho \geq a) = \sum_{n=-\infty}^{\infty} c_n \Psi_n(\rho, \varphi) \ , \qquad \text{(2-33b)}$$

where the Fourier coefficients, b_n and c_n, can be determined by enforcing boundary conditions at the cylinder wall. These boundary conditions take the forms

$$E^i + E^s(\rho \geq a) = E^s(\rho \leq a), \qquad \text{(2-34a)}$$

$$H_\varphi^i + H_\varphi^s(\rho \geq a) = H_\varphi^s(\rho \leq a). \qquad \text{(2-34b)}$$

The φ component of the H field can be obtained by assuming an $e^{j\omega t}$ time variation and using Maxwell's equation given in (2-9a) as

$$H_\varphi = \frac{-j}{\omega\mu} \frac{\partial E}{\partial \rho} \ . \qquad \text{(2-35)}$$

The incident field can also be expressed as an infinite series by using the addition theorem for Hankel functions [30]. The Fourier coefficients can now be found by using (2-34a) and (2-34b) then formally differentiating through the infinite summation as suggested in (2-35). After some manipulation the following result is obtained for the scattered field exterior to the cylinder [31],

$$E^s(\rho \geq a) = \sum_{n=-\infty}^{\infty} d_n H_n^{(2)}(-j\gamma_e \rho) \ H_n^{(2)}(-j\gamma_e \rho_t), \qquad \text{(2-36)}$$

where

$$d_n = \begin{cases} K_0/Q_0, & \text{for } n=0 \\ 2K_n \cos n(\varphi-\varphi_t)/Q_n, & \text{for } n>0 \end{cases} \tag{2-37}$$

and

$$K_n = \gamma_a a \; J_n(-j\gamma_e a)\; J_{n+1}(-j\gamma_a a) -$$

$$+ \gamma_e a \; J_n(-j\gamma_a a)\; J_{n+1}(-j\gamma_e a), \tag{2-38}$$

$$Q_n = \gamma_a a \; H_n^{(2)}(-j\gamma_e a)\; J_{n+1}(-j\gamma_a a) -$$

$$+ \gamma_e a \; J_n(-j\gamma_a a)\; H_{n+1}^{(2)}(-j\gamma_e a). \tag{2-39}$$

Of importance in using (2-36), is the number of terms needed in the summation to get accurate results. In order to determine the number of terms needed, we first make use of the asymptotic expansion of the Hankel function,

$$H_n^{(2)}(z) \quad \doteq \quad \left[\frac{2j}{\pi z}\right]^{1/2} j^n e^{-jz}, \tag{2-40}$$

which is valid for large $|z|$. From this expansion, it can be seen that for the antenna and field point far from the cylinder axis (i.e., large z), the Hankel functions in the summation do not increase in magnitude with increasing order. Therefore, it is only necessary to evaluate the behavior of d_n with increasing n. Since d_n is a function of the external propagation constant, the cylinder radius, and the order, its behavior will depend on these parameters. To show some typical responses of d_n with increasing order, in Fig. 2.12 $|d_n|$ is plotted versus n for radii of 1 and 2 meters. For both of these cases the magnitude of d_n is insignificant for n>15, indicating that less than 15 terms in the summation need to be evaluated. In evaluating the Bessel and Hankel functions needed for this figure, the FORTRAN subroutines described in [33] were used.

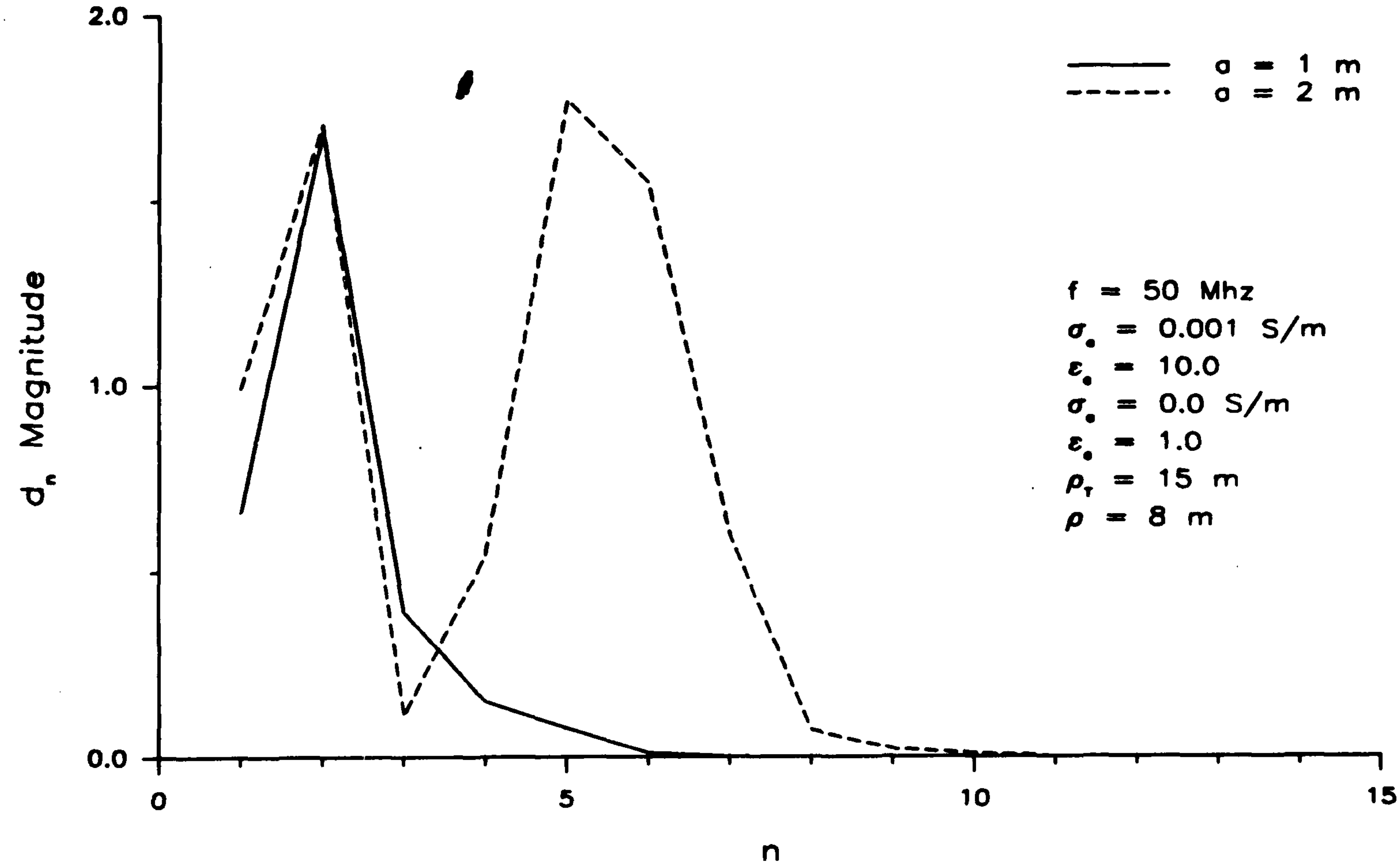


Fig. 2.12. Plot of the magnitude of the coefficient term in the eigenfunction solution for the scattered field versus integer order, for cylinder radii of 1 and 2 meters.

In summary, a method for calculating the electromagnetic field generated by a line source antenna in the earth and scattered by a circular cylinder whose axis is parallel to the line source is given by (2-36). Although there are times when this type of modeling is important (e.g., in finding the electromagnetic response of a circular tunnel), a more general modeling technique is desirable. Such a technique is described below.

2.3.5 Sinusoidal Response Using the Volume Current Method

In this section the anomalies are assumed to be cylinders of arbitrary cross sectional shape with axes parallel to the line source antenna. An approximate solution to finding the scattered electromagnetic field can be achieved by representing the anomalies by small circular cylinders [34], [35]. The total field is then found as the sum of the contribution from each cylinder. This technique for solving for the scattered electromagnetic field is commonly referred to as the volume current method (VCM), it is from the class of numerical techniques called moment methods [36]. A typical geometry for using this method is shown in Fig. 2.13. The figure illustrates how a large circular cylinder would be approximated using this technique.

Reference [35] gives a good description of the procedure to follow in generating electromagnetic cross-borehole data using VCM, and an outline of this procedure is in Appendix A. A couple of items concerning the use of VCM need to be mentioned:

a) As suggested by the authors, using 8 - 10 small cylinders per intrinsic wavelength gives results which compare favorably with the exact solution (i.e., for a circular cylinder).

b) In cases where the anomalous cylinder(s) has zero conductivity (e.g., an air filled tunnel), the method appears to break down. The problem is that with zero conductivity, the equivalent currents cannot be sustained in the anomaly. To resolve this

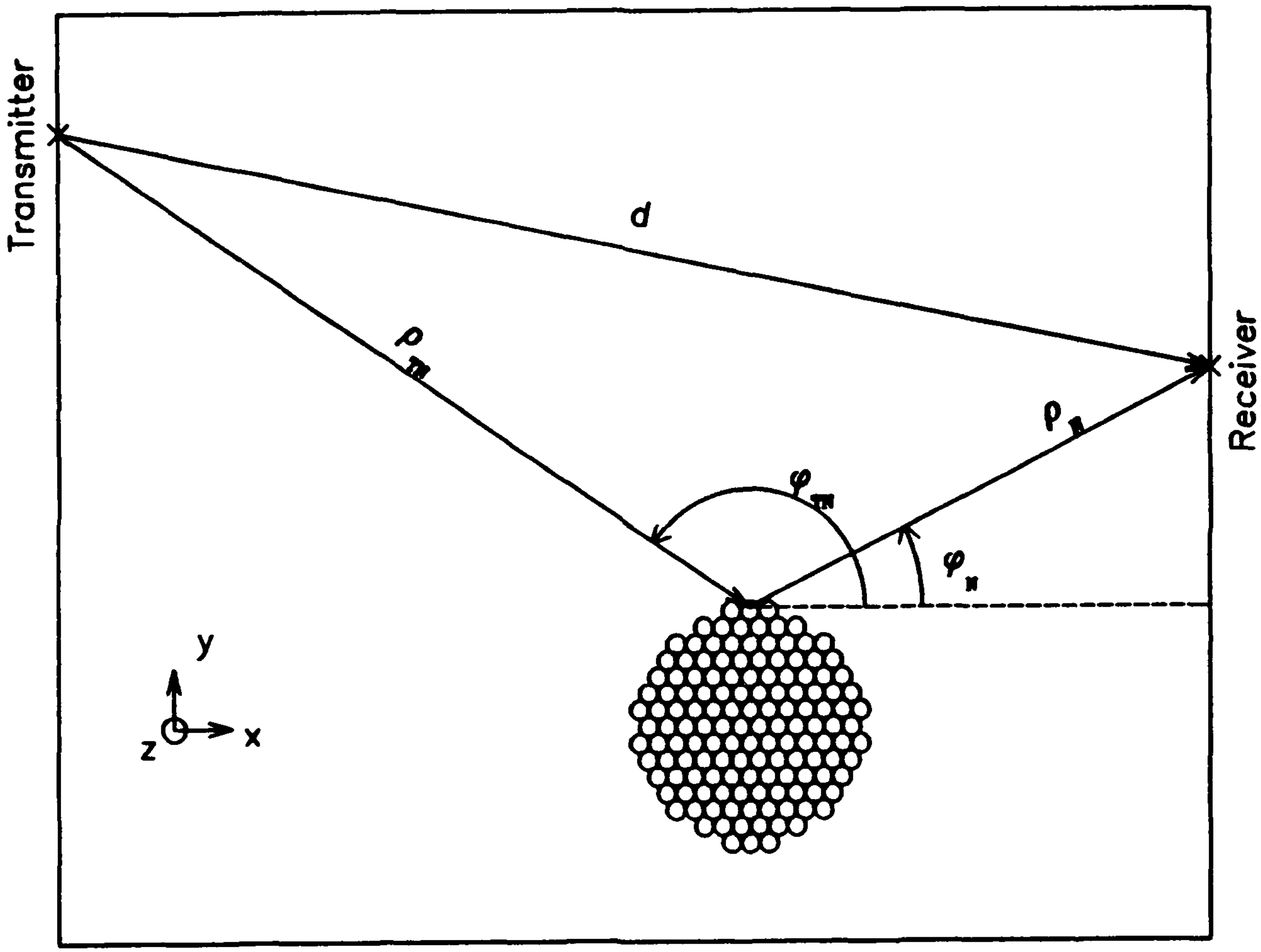


Fig. 2.13. Typical configuration for VCM solution of cross-borehole simulation problem.

problem, the tunnel is modeled as having a small conductivity (less than 0.0001 S/m), and the result is then checked against the exact solution for a circular tunnel.

Such a comparison between the VCM solution and the exact solution is shown in Fig. 2.14. As can be seen from the figure the agreement is quite good.

2.3.6 Sinusoidal Response Using the Born Approximation

We now consider solving the problem of finding the sinusoidal response of a homogeneous earth containing an anomalous region using an approximation which is valid when the electrical parameters of the region are similar to the surrounding earth. This approximation is important since it is the basis for a reconstruction method which will be discussed in Chapter III. The method to be used is described in [37]. Again, the total electric field is considered to be the sum of incident and scattered fields as in (2-31). Assuming that the field is to be calculated at locations away from the source, equation (2-23) reduces to,

$$(\nabla^2 - \gamma^2)E(\mathbf{\Sigma}) = 0, \qquad (2\text{-}41)$$

where γ is a function of position, and the electric field's dependence on position is explicitly noted. Since isolated anomalies are being considered, the propagation constant can be represented as,

$$\gamma(\mathbf{\Sigma}) = \gamma_e n(\mathbf{\Sigma}), \qquad (2\text{-}42)$$

where γ_e is the propagation constant of the external medium, and $n(\mathbf{\Sigma})$ represents a perturbation from the background propagation constant. Equation (2-41) can be re-written using (2-42) as

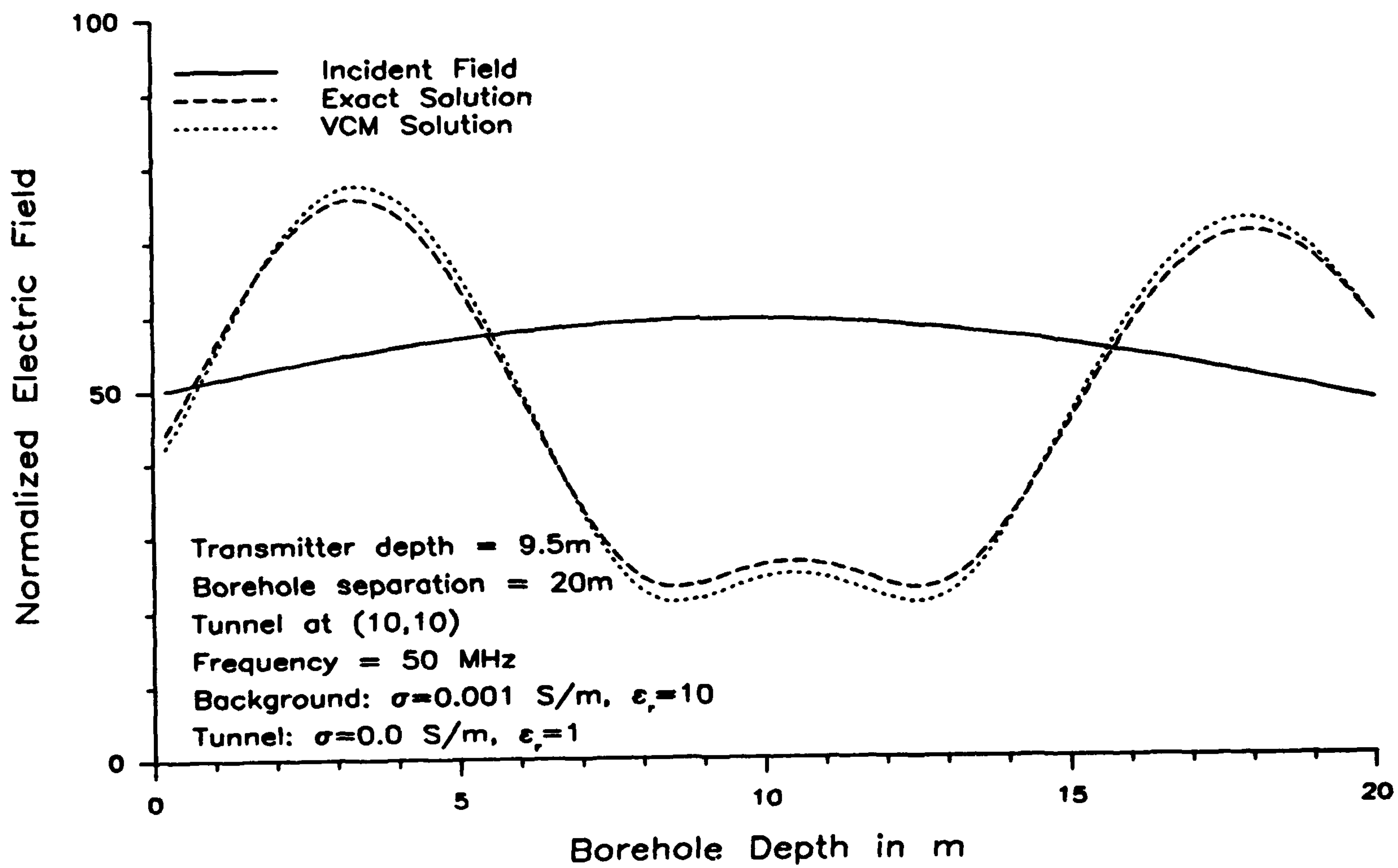


Fig. 2.14. Received electric field magnitude versus borehole depth using exact solution and VCM.

$$\{\nabla^2 - \gamma_e^2[1-1+n^2(\mathbf{x})]\}E(\mathbf{x}) = 0, \tag{2-43a}$$

$$(\nabla^2 - \gamma_e^2)E(\mathbf{x}) = \gamma_e^2[n^2(\mathbf{x})-1]E(\mathbf{x}), \tag{2-43b}$$

$$(\nabla^2 - \gamma_e^2)[E^i(\mathbf{x})+E^s(\mathbf{x})] = \gamma_e^2[n^2(\mathbf{x})-1]E(\mathbf{x}), \tag{2-43c}$$

$$(\nabla^2 - \gamma_e^2)E^s(\mathbf{x}) = \gamma_e^2[n^2(\mathbf{x})-1]E(\mathbf{x}), \tag{2-43d}$$

where we have used the fact that the incident field solves the wave equation in the external medium, that is,

$$(\nabla^2 - \gamma_e^2)E^i(\mathbf{x}) = 0, \tag{2-44}$$

Now, (2-43d) can be considered to be the wave equation for the scattered field with the source term as shown on the right hand side of the equation. This equation can be solved by finding the Green's function solution to

$$(\nabla^2 - \gamma_e^2)G(\mathbf{x}) = \delta(\mathbf{x}), \tag{2-45}$$

and then by convolving $G(\mathbf{x})$ with the source term as

$$E^s(\mathbf{x}) = \int_A\int \gamma_e^2[n^2(\mathbf{x}')- 1]E(\mathbf{x}')G(|\mathbf{x}-\mathbf{x}'|)d\mathbf{x}' \tag{2-46}$$

where the integration is over the area of the anomalies. The Green's function can be found by referring to (2-30) as

$$G(|\mathbf{x}-\mathbf{x}'|) = \frac{-j}{4} H_0^{(2)}(-j\gamma_e|\mathbf{x}-\mathbf{x}'|). \tag{2-47}$$

The first Born approximation involves neglecting the contribution

of the scattered field in the right hand side of (2-46) to get

$$E^{s}(\mathbf{x}) \simeq \int_{A}\!\!\int \gamma_e^2[n^2(\mathbf{x}')-1]E^{i}(\mathbf{x}')G(|\mathbf{x}-\mathbf{x}'|)d\mathbf{x}'. \qquad (2\text{-}48)$$

This equation was used to find the approximate electromagnetic response of a low conductivity circular cylinder buried in the earth. The result of using the Born approximation is plotted along with the exact solution in Fig. 2.15. The characteristics of the background earth are the same as in Fig. 2.14. The cylinder represents a slight inhomogeneity in that its conductivity (σ=0.0005 S/m) and permittivity (ε_r=8) are only mildly different than the background. The anomaly's characteristics can not be much different than the background in order to justify dropping the contribution from the scattered field in (2-46).

Although there is good agreement between the Born approximation and the exact solution in the figure, as stated above the Born approximation is only useful for determining the response of mildly scattering objects. The main interest in studying the Born approximation is that it and the Rytov approximation are used in the Fourier diffraction theorem [37] which will be considered in the next chapter as a means of image reconstruction.

2.4 Transient Response of a Homogeneous Earth Containing Isolated Anomalies

2.4.1 Introduction

In this section we consider solving the problem of finding the time signal at a receiver location given the input time signal at a transmitter location. This problem is important for characterizing the time-of-flight measurement process used in geotomography. In this

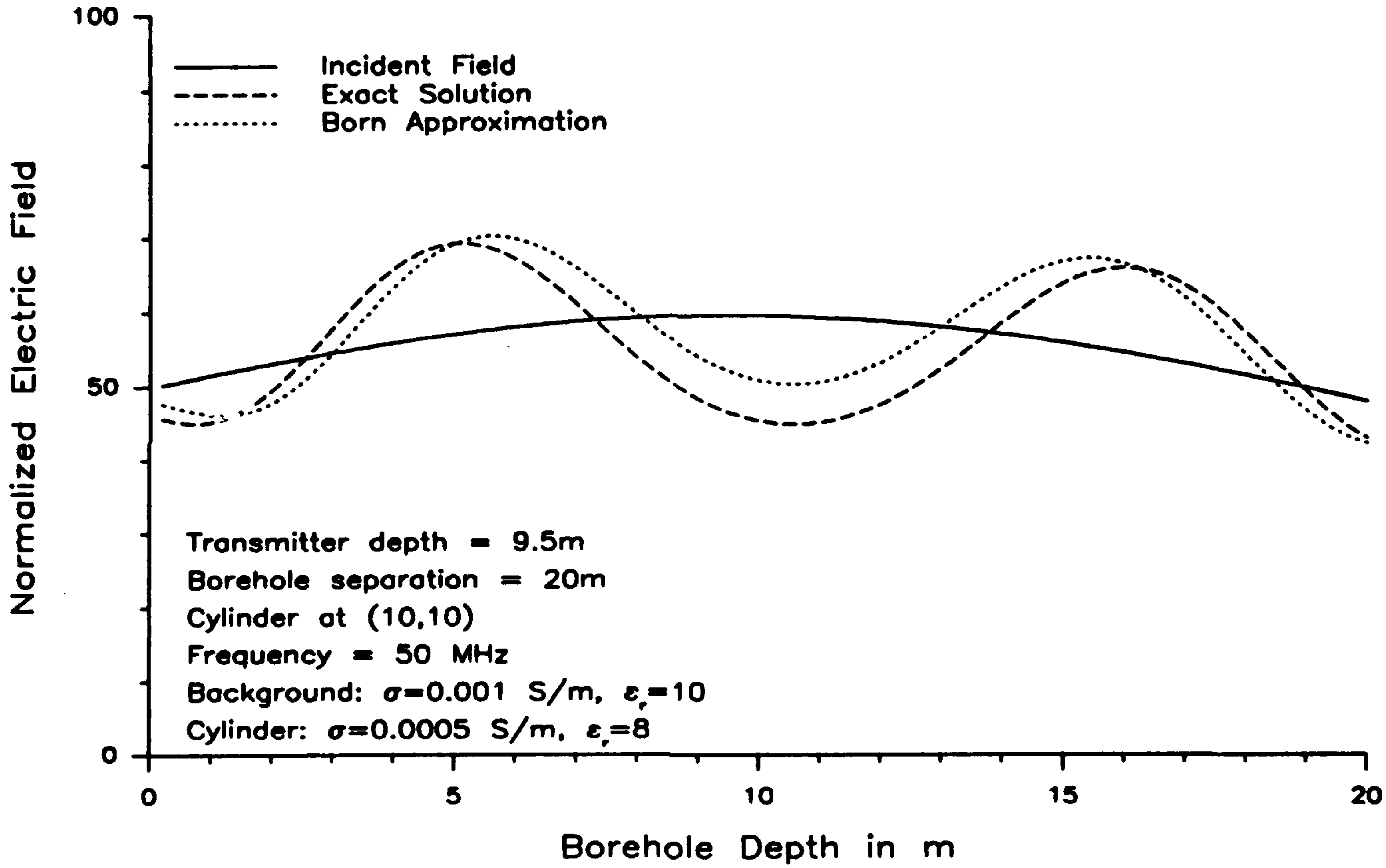


Fig. 2.15. Received electric field magnitude versus borehole depth using exact solution and Born approximation.

process the time it takes a pulse to travel from transmitter to receiver is measured, and in this way the intrinsic velocity of the medium can be obtained. The results in this section draw upon the those obtained in the last section. We first consider using the Green's function solution for a homogeneous earth.

2.4.2 Time Response of a Homogeneous Earth Using the Green's Function Solution

The Green's function solution for a line source antenna in a homogeneous earth was derived in the last section, this function is plotted against time in Fig. 2.16. It is useful for finding the electromagnetic response to arbitrary time functions. For example, in the GPEMS cross-hole radar system [38], one cycle of a 100 MHz sine wave is used as a transmission signal. This means that the source term will take the form

$$J = \sin\omega t[H(t) - H(t - \frac{2\pi}{\omega})], \tag{2-49}$$

with $\omega = 2\pi \times 100 \times 10^6$. With this input function the right hand side of the damped wave equation becomes

$$\mu\frac{\partial J}{\partial t} = \mu\omega\cos\omega t[H(t) - H(t - \frac{2\pi}{\omega})] + \mu\sin\omega t[\delta(t) - \delta(t - \frac{2\pi}{\omega})]. \tag{2-50}$$

The electromagnetic response to such an input can be found by convolution, as

$$E(\rho,t) = -G(\rho,t) * \mu\frac{\partial J}{\partial t}. \tag{2-51}$$

Fig. 2.17 is the result of carrying out this operation numerically.

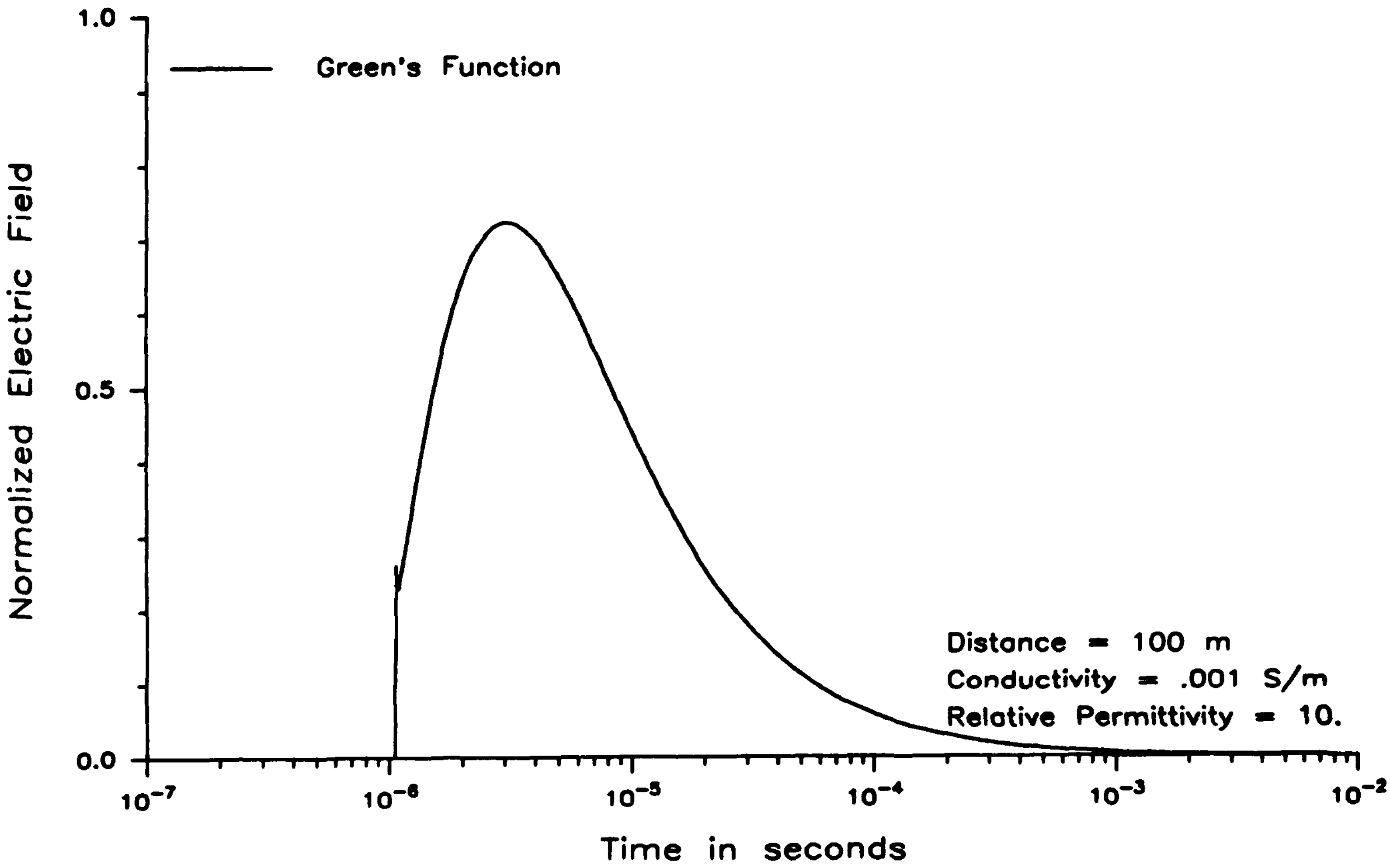


Fig. 2.16. Green's function for the damped wave equation plotted against time for a radial distance of 100 m from the line source antenna.

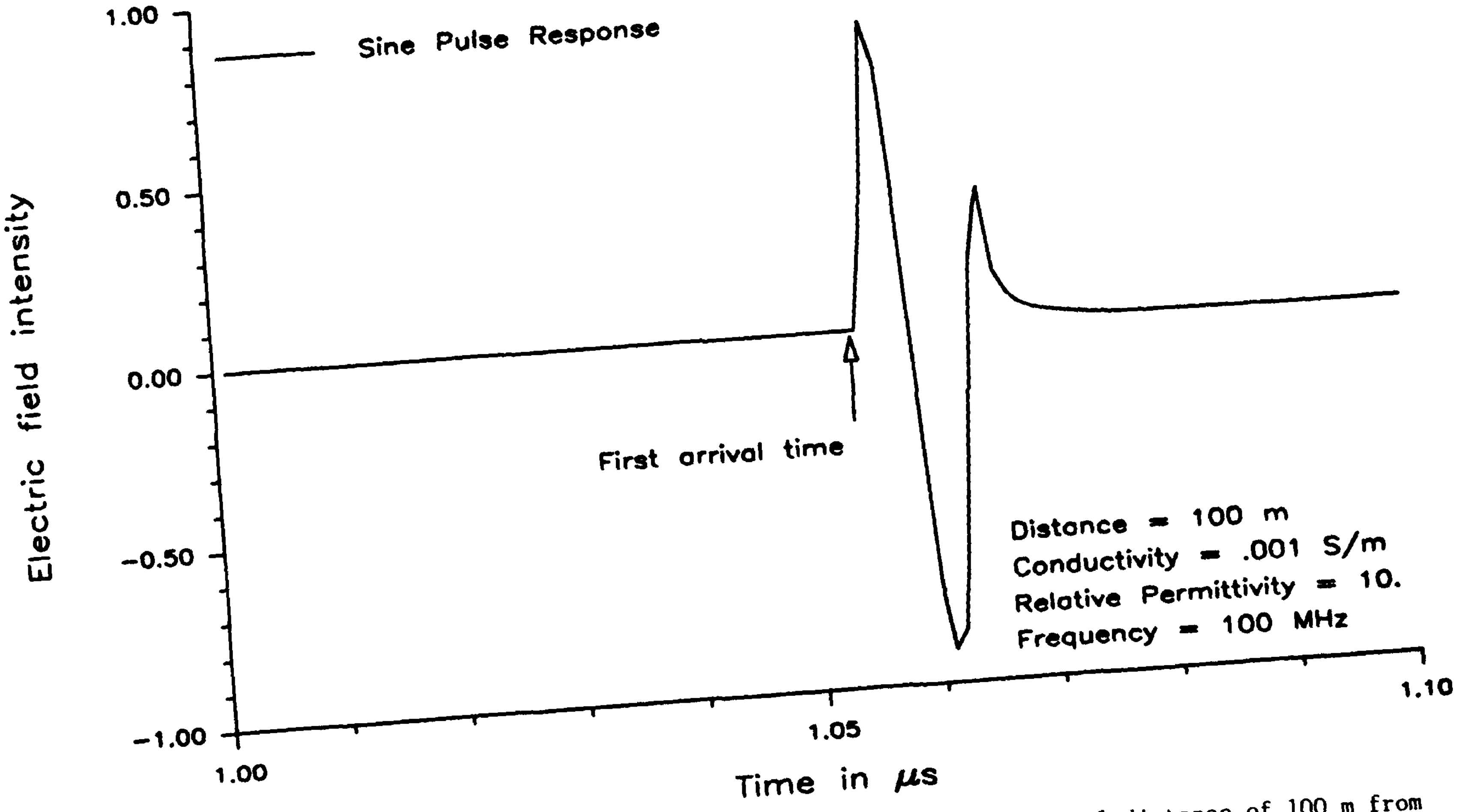


Fig. 2.17. 100 MHz sine pulse response for a homogeneous earth at a radial distance of 100 m from the line source antenna

Note that the first arrival time is at approximately 1.055 μs, which is the time that would be obtained by dividing the distance by the zero conductivity velocity. Also note that the time of arrival of the first peak (the quantity which would be measured in a TOF system) is approximately 0.0025 μs (1/4 cycle) after the first arrival time. This is to be contrasted with the result for a 1 MHz sine pulse shown in Fig. 2.18, where the first peak arrives about 30% slower than 1/4 cycle of the source signal. This discrepancy can be resolved by referring again to Fig. 2.7, and noting that although the velocity assuming zero conductivity matches the true velocity at 100 MHz, this is not the case at 1 MHz.

2.4.3 Fourier Transform Method for Finding the Transient Response

The method described above can not be used for finding the transient response of a cross section of earth containing arbitrarily shaped anomalies. However, we can use the results of the last section in conjunction with Fourier analysis in order to solve this problem. This allows the calculation of the time response of the electromagnetic field at the receiving antenna given the input signal at the transmitting antenna. The process involves the following steps:

a) Calculate the sinusoidal response of the earth containing isolated anomalies over a range of frequency values starting at frequencies close to zero.

b) Fourier transform the input signal (e.g., the sine pulse mentioned previously) either analytically or using the fast Fourier transform (FFT) algorithm.

c) Multiply the results of a) and b) together for each frequency value selected.

d) Inverse transform the result of c) using the FFT algorithm.

Fig. 2.19 shows the responses of a homogeneous earth and a homogeneous earth containing a tunnel. Two items of interest in this figure are:

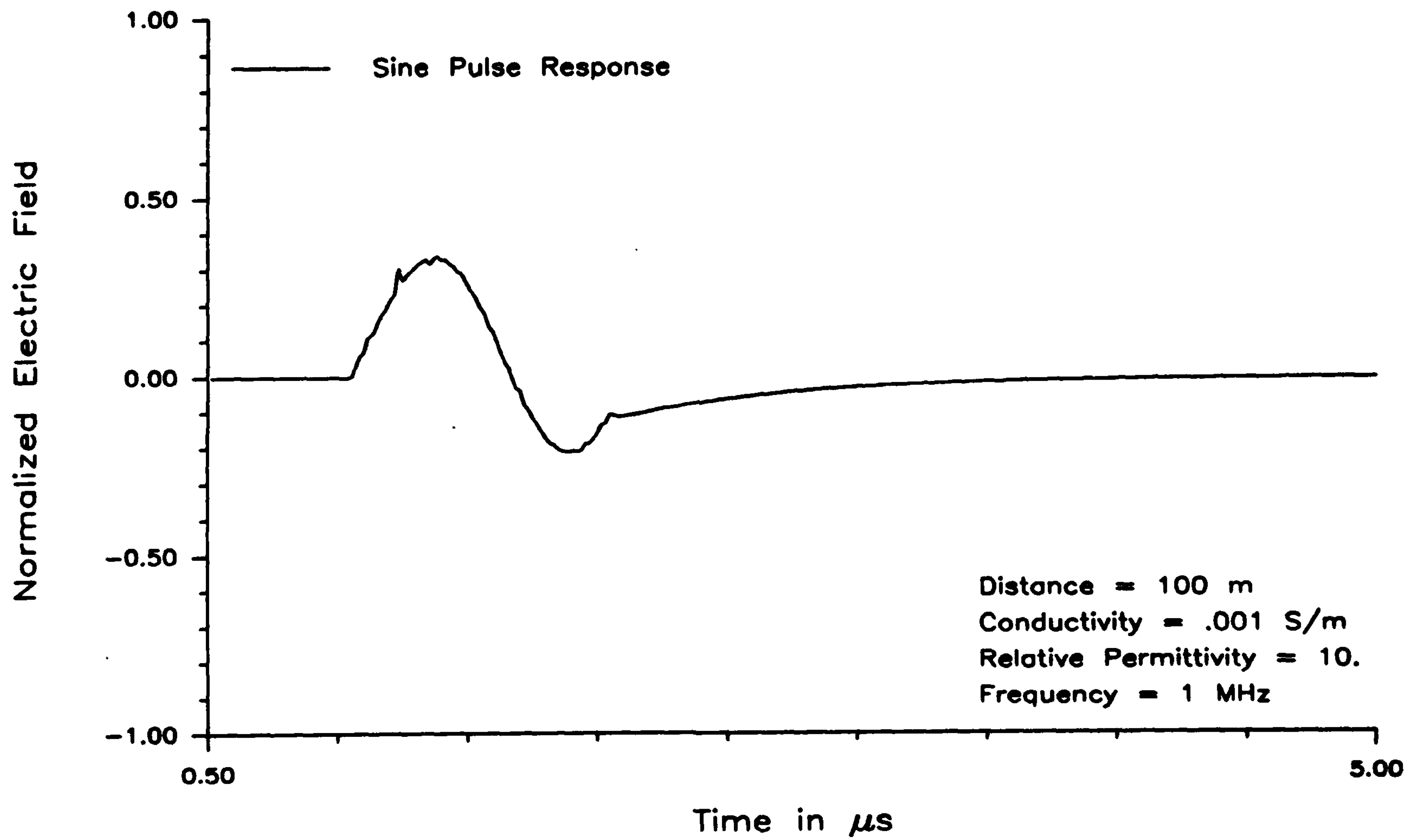


Fig. 2.18. 1 MHz sine pulse response for a homogeneous earth at a radial distance of 100 m from the line source antenna.

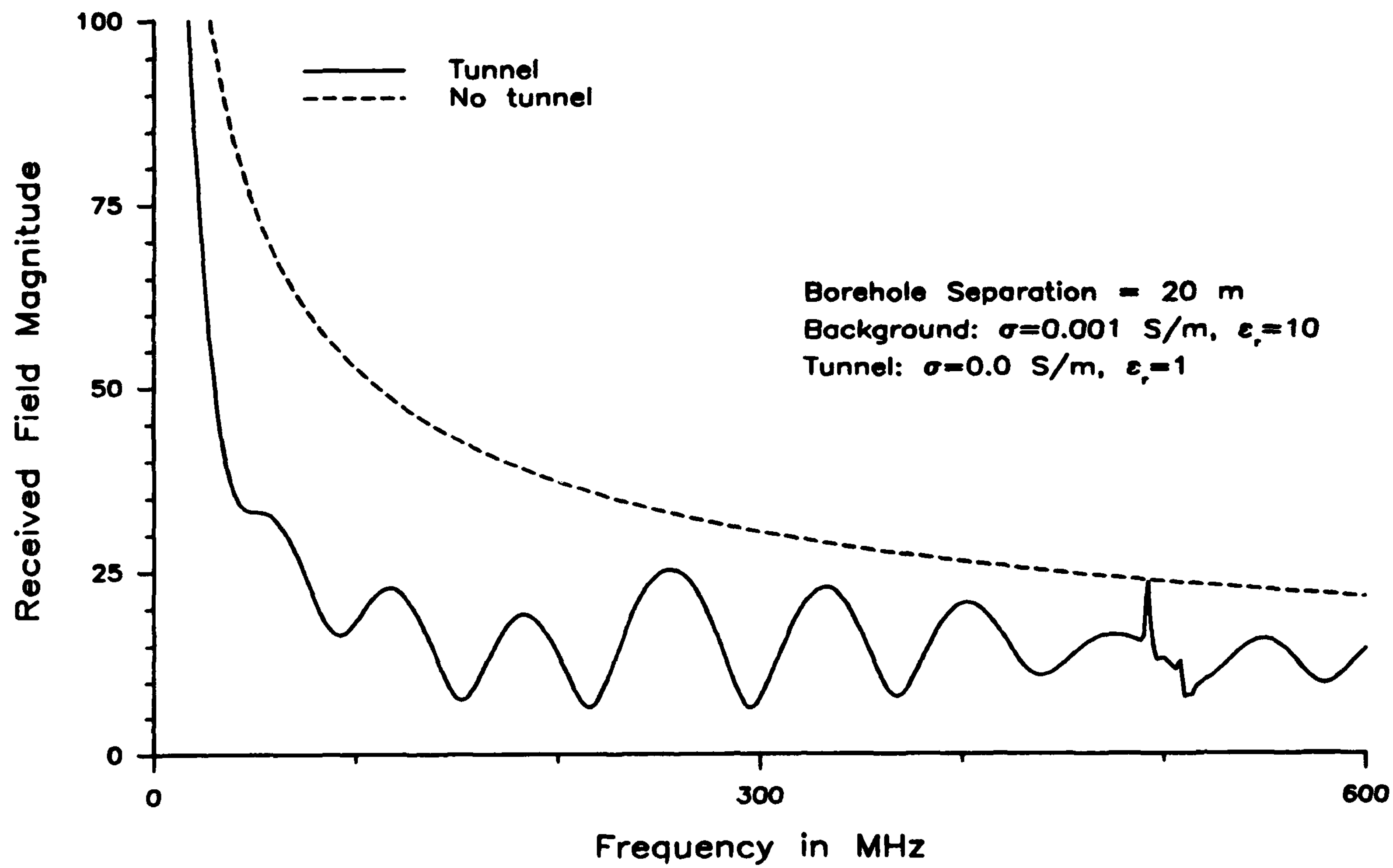


Fig. 2.19. Magnitude response versus frequency for homogeneous earth, and earth containing a tunnel.

a) Although the magnitude response falls off rapidly at low frequencies, it decays slowly above 100 MHz. Of course, this behavior could have been predicted by referring to Fig. 2.1.

b) Although the homogeneous response is relatively smooth above 100 MHz, with the tunnel included, noticeable oscillations occur in the response. These oscillations are sometimes referred to as resonances, and are the result of reflections at the tunnel wall causing cancellation and reinforcement of the transmitted waves.

The spike in the plot for the tunnel at about 500 MHz is the result of a numerical difficulty in calculating the complex Bessel functions at this frequency. This does not seem to cause any errors in the results that follow.

For the examples that follow, the input signal will be the sine pulse of (2-49). Using the convolution theorem, its Fourier transform

$$S(\omega) = \frac{\tau}{j\sqrt{2\pi}}\left[e^{j\tau(\omega_c-\omega)/2}\,\frac{\sin\tau(\omega_c-\omega)/2}{\tau(\omega_c-\omega)} - e^{-j\tau(\omega_c+\omega)/2}\,\frac{\sin\tau(\omega_c+\omega)/2}{\tau(\omega_c+\omega)}\right], \tag{2-52}$$

where τ (= $2\pi/\omega_c$) is the pulse duration time. We need only to multiply this transform by the frequency response, and then inverse transform to obtain the time signal at the receiving antenna. This operation was performed for both the homogeneous earth, and the earth containing the tunnel. The time responses are shown in Figs. 2.20 and 2.21. As can be seen from the second figure, the tunnel causes multiple pulses to reach the receiver. The first pulse seen is the result of the electromagnetic ray traveling through the tunnel to reach the receiver. This ray has the shortest travel time, since part of the distance travelled is through air, which has a higher wave velocity than the earth. After the first pulse arrives there is the interference between this pulse and a pulse which has twice reflected inside the tunnel and

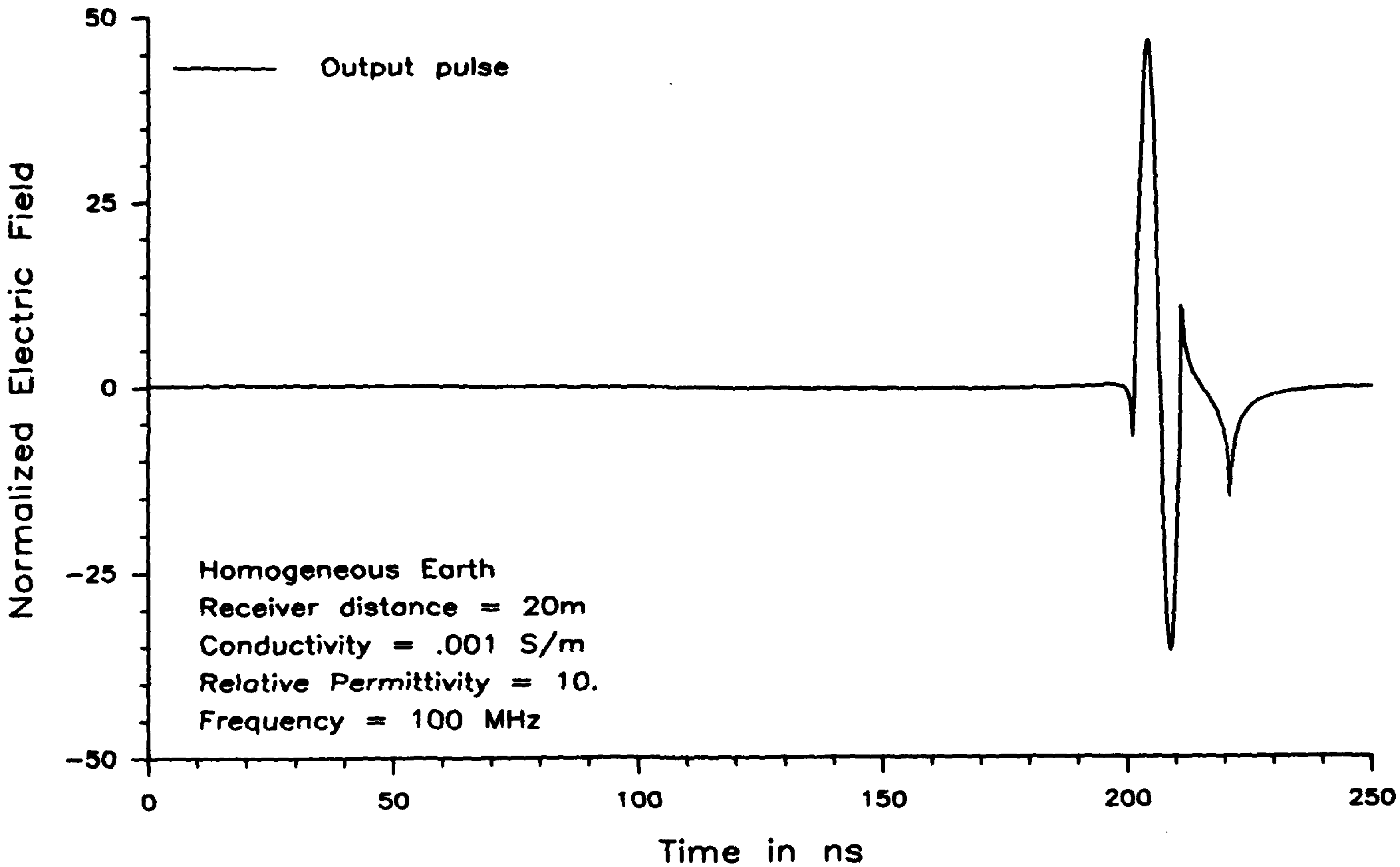


Fig. 2.20. Transient response of a *homogeneous* earth *for* a 100 MHz sine pulse input.

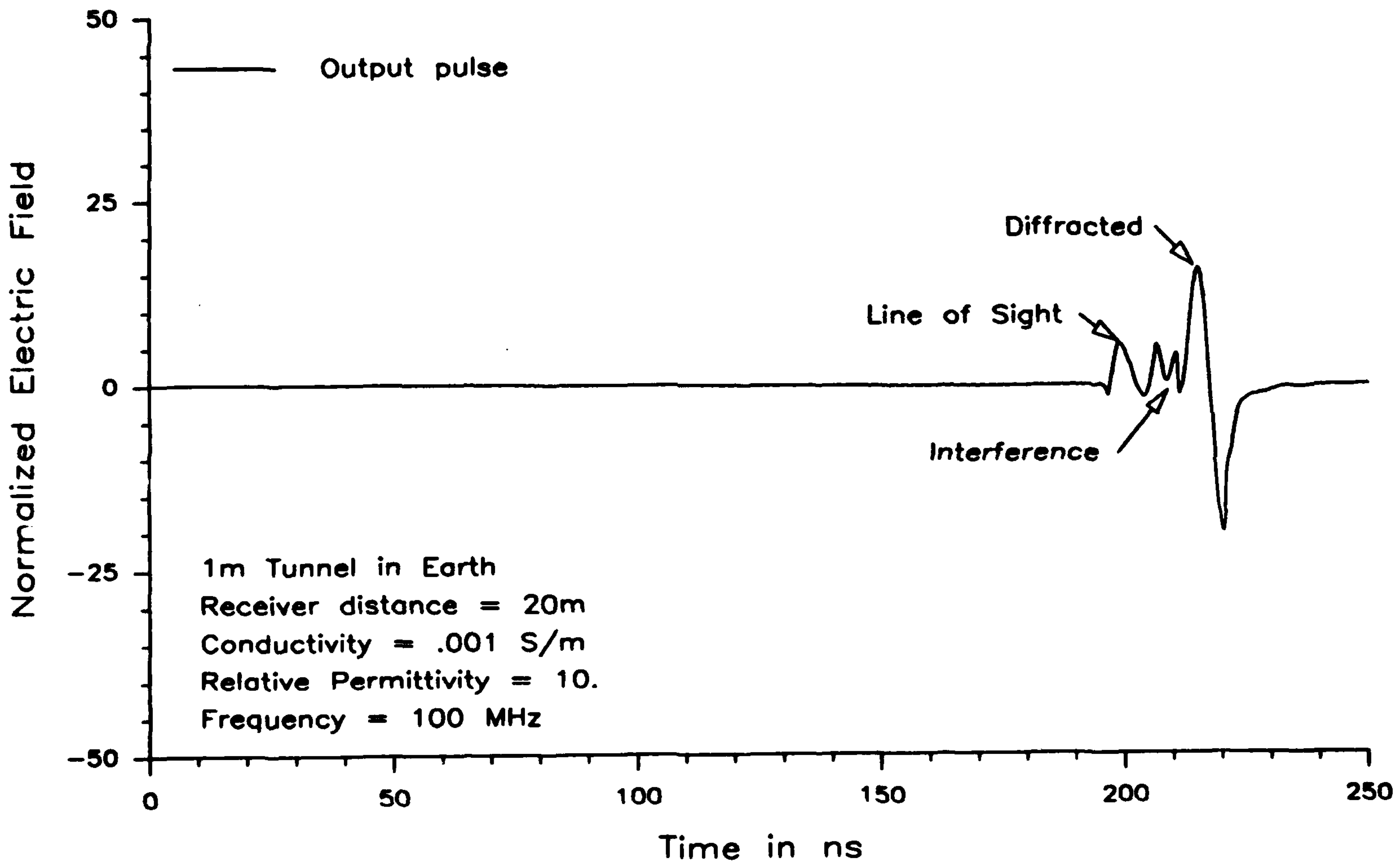


Fig. 2.21. Transient response of a homogeneous earth containing a tunnel for a 100 MHz sine pulse input.

then is detected at the receiver. Finally, we see the pulse which has diffracted from the tunnel ceiling and floor before reaching the receiver. Surprisingly, this pulse has the largest amplitude.

Although this frequency domain method is of use for predicting the transient response of a tunnel in the earth or any other buried circular cylindrical object, it becomes impractical for arbitrarily shaped objects in that it is very time consuming to use the volume current method over a wide range of frequencies. In addition, the frequency domain method only gives us information at a single receiver location, and it might be useful to obtain the response over the entire cross sectional region between the boreholes. For these and other reasons, a ray optics approach is investigated in the next section.

2.5 Ray Optics Method

2.5.1 Introduction

In the ray optics approach the assumption is made that in a homogeneous medium an electromagnetic wave follows straight paths from the transmitter to the receiver. For heterogeneous media we still consider the electromagnetic wave to follow straight paths although the ray may be subject to refractions, reflections, and/or diffractions at boundaries between different media. These effects will alter the direction of the ray.

The value of the electric field can be determined at any point along its ray path by using (2-30) and the large argument asymptotic expansion for the Hankel function to get

$$E = \frac{-\omega\mu I}{4} H_0^{(2)}(-j\gamma\rho) \sim \frac{-\omega\mu I}{4} \left[\frac{-2}{\pi\gamma\rho}\right]^{1/2} e^{-\gamma\rho}, \qquad (2\text{-}53)$$

where the electric field is measured at a radial distance, ρ, from the

line source transmitter. In order to find the electric field at a point in a heterogeneous medium, we need to find the ray linking the transmitter and the receiver after taking into account the effects of refraction, reflection, and diffraction. These effects will be discussed in subsequent sections. It is worth mentioning that in heterogeneous media there will often be more than one ray path linking transmitter and receiver. If such is the case, the electric field should be calculated by summing the contributions from the different paths.

Before discussing the details of ray optics, its importance in the geotomography setting needs to be highlighted. First of all, the calculated fields using ray optics will not be as accurate as those obtained using the methods described above. Also, although it is often true that the ray optics method is more computationally efficient than the VCM, it is often much more difficult (i.e., time consuming) to program, thereby nullifying any net computational advantage.

Rather, the main reason for considering the ray optics approach is that in performing the inversion of cross-hole data (i.e., geotomography), the assumption is often made that the electromagnetic wave follows ray paths between transmitters and receivers. Therefore, two immediate reasons for studying ray optics are to address the following questions.

a) Under what conditions is the straight ray path a good approximation, and

b) If it is not a good approximation, is there anything that can be done to improve the approximation?

With these questions in mind we consider some of the details involved in the ray optics method.

2.5.2 Refraction and Reflection of Electromagnetic Waves in Lossy Media

Consider a ray obliquely incident on a boundary between two lossy

media. The configuration is shown in Fig. 2.22. The incident, reflected, and transmitted rays are assumed to lie in the x-y plane, making the problem two-dimensional. The electric fields along the three rays can be expressed as

$$\hat{E}_i = \hat{a}_z E_i e^{-j\beta_1 \xi_1}, \tag{2-54a}$$

$$\hat{E}_r = \hat{a}_z E_r e^{-j\beta_1 \xi_2}, \text{ and} \tag{2-54b}$$

$$\hat{E}_t = \hat{a}_z E_t e^{-j\beta_2 \xi_3}, \tag{2-54c}$$

where

$$E_i = E\, e^{-\alpha_1 \xi_1'}, \tag{2-55a}$$

$$\xi_1 = y\sin\varsigma_1 + x\cos\varsigma_1, \tag{2-55b}$$

$$\xi_1' = y\sin[\varsigma_1 + \rho_1] + x\cos[\varsigma_1 + \rho_1], \tag{2-55c}$$

$$E_r = \Gamma E e^{\alpha_1 \xi_2'}, \tag{2-55d}$$

$$\xi_2 = -y\sin\varsigma_1 + x\cos\varsigma_1, \tag{2-55e}$$

$$\xi_2' = -y\sin[\varsigma_1 + \rho_1] + x\cos[\varsigma_1 + \rho_1], \tag{2-55f}$$

$$E_t = TEe^{\alpha_2 \xi_3'}, \tag{2-55g}$$

$$\xi_3 = y\sin\varsigma_2 + x\cos\varsigma_2, \text{ and} \tag{2-55h}$$

$$\xi_3' = y\sin[\varsigma_2 + \rho_2] + x\cos[\varsigma_2 + \rho_2]. \tag{2-55i}$$

Γ and T are the reflection and transmission coefficients (factors), and E is the value of the incident field at the interface. For the incident ray, the direction of travel is along the ξ_1 direction, but the ray attenuates along the ξ_1' directions. The reflected and transmitted (refracted) rays have similar directions of travel and attenuation. As can be seen from the figure, the incident ray is at an angle of ς_1 from

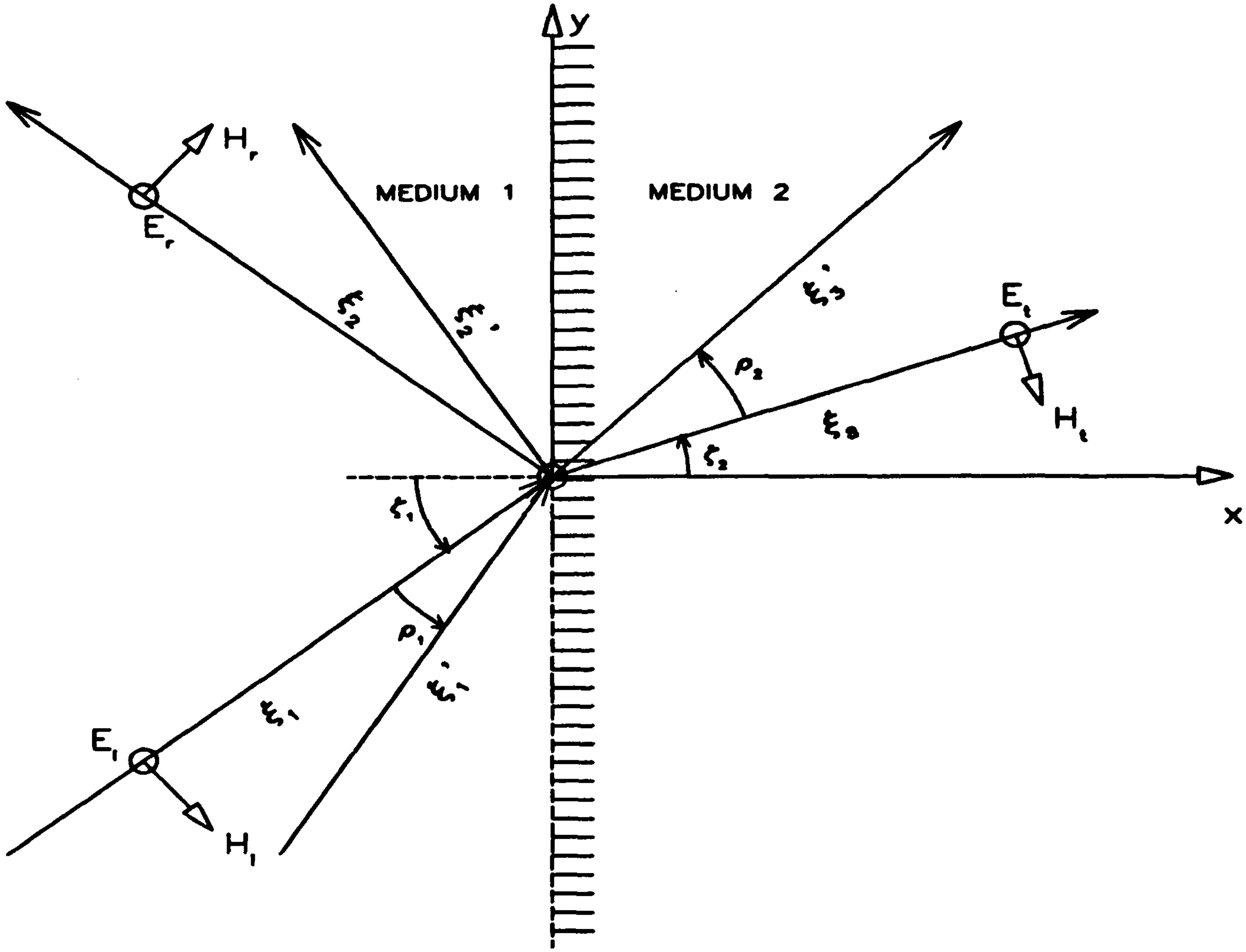


Fig. 2.22. Illustration of an incident plane wave being reflected and refracted at an interface between two lossy media.

the normal to the interface. If the incident wave is nonuniform (i.e., the direction of attenuation differs from the direction of travel), then $\zeta_1+\rho_1$ is the angle from the normal at which the wave attenuates. If the wave is uniform, then ρ_1 is zero. In this development it is assumed that the wave starts out from the antenna being uniform, and then when it encounters an interface it becomes nonuniform as predicted by the equations above.

For an incident plane wave, the relations between the incident and transmitted directions were derived in [39]. These relations are,

$$\alpha_1 \sin(\zeta_1+\rho_1) = \alpha_2 \sin(\zeta_2+\rho_2), \text{ and} \tag{2-56a}$$

$$\beta_1 \sin\zeta_1 = \beta_2 \sin\zeta_2 \ , \tag{2-56b}$$

where α_2 and β_2 are different from the intrinsic constants (α_{02}, β_{02}) in medium 2, and are found from,

$$2\beta_2^2 = |\gamma_1|^2 - \mathrm{Re}(\gamma_{02}^2) + |\gamma_1^2 - \gamma_{02}^2|, \text{ and} \tag{2-57a}$$

$$2\alpha_2^2 = |\gamma_1|^2 + \mathrm{Re}(\gamma_{02}^2) + |\gamma_1^2 - \gamma_{02}^2| \ . \tag{2-57b}$$

The symbol γ_{02} in the equations above is the intrinsic propagation constant in medium 2. For the configuration being considered, the waves are cylindrical since the transmitting antenna is a line source. But, if the interface is assumed to be far from the antenna, then the waves at the interface are locally plane waves, so that in this far field case, the above equations can be used to find the reflected and refracted rays.

The transmission and reflection factors were not derived in [39], but are easily found by requiring the sum of the incident and reflected electric and magnetic fields to be equal to the transmitted electric and magnetic fields at the interface. This requirement results in,

$$\Gamma = \frac{\eta_{yt} - \eta_{yi}}{\eta_{yi} + \eta_{yt}} \ , \tag{2-58a}$$

$$T = \frac{2\eta_{yt}}{\eta_{yi} + \eta_{yt}} , \tag{2-58b}$$

$$\eta_{yt} = \frac{-j\omega\mu}{\alpha_2 \cos(\xi_2+\rho_2) + j\beta_2 \cos\xi_2} , \text{ and} \tag{2-58c}$$

$$\eta_{yi} = \frac{-j\omega\mu}{\alpha_1 \cos(\xi_1+\rho_1) + j\beta_1 \cos\xi_1} . \tag{2-58d}$$

Fig. 2.23 illustrates how this theory can be used to determine the field at a receiver using the ray optics approach. The ray travels a distance ρ_1, and reaches the first interface. The transmission coefficient at this interface can be calculated from the equations given above. The propagation constant times the distance along this first path will be defined by

$$\gamma_{d1} := \gamma_1 d_1 . \tag{2-59}$$

However, over ρ_2, it is defined by

$$\gamma_{d2} := \alpha_2 d_2 \cos(\rho_2) + j\beta_2 d_2 , \tag{2-60}$$

where ρ_2 is the angle between the phase (direction of travel) and attenuation directions of the transmitted rays. Similarly, for the ray along d_3, its propagation constant times distance traveled is defined by

$$\gamma_{d3} := \alpha_3 d_3 \cos(\rho_3) + j\beta_3 d_3 , \tag{2-61}$$

Finally, by summing all length contributions, the electric field at the receiver is given by , see also (2-40),

$$E_{rcv} = \frac{-j\omega\mu I}{4}\left[\frac{2}{\pi}\right]^{1/2} \frac{e^{-(\gamma_{d1}+\gamma_{d2}+\gamma_{d3})}}{\sqrt{\gamma_{d1}+\gamma_{d2}+\gamma_{d3}}} \text{ T1 T2} \tag{2-62}$$

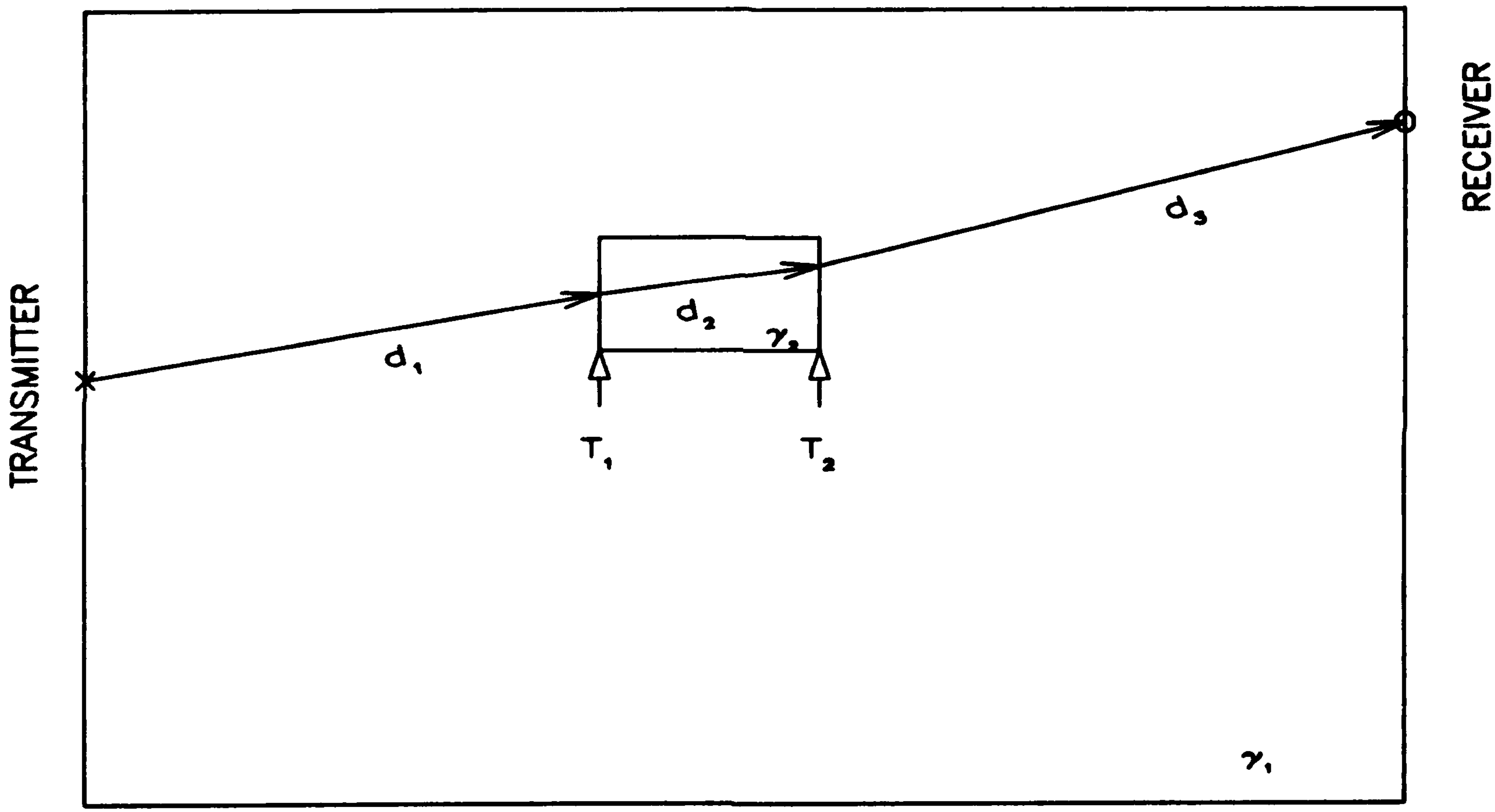


Fig. 2.23. Illustration of a ray which refracts at two interfaces before reaching the receiver.

where T1 and T2 are the transmission coefficients at the first and second interfaces. A similar formulation can be made for a ray which reflects from an interface at an oblique angle, and then reaches the receiver. In this reflection case, T1 or T2 in equation (2-62) would then be replaced by a reflection coefficient.

Even in this two interface example presented here, it is not difficult to see that the process of linking a transmitter with a receiver over a path which involves refractions and/or reflections will not be a trivial matter.

2.5.3 Diffraction of Electromagnetic Waves from a Lossy Wedge

Diffraction theory was developed as an approximate high frequency technique to account for the fields which are present in the shadow region of a conducting object [40], [41]. For a depiction of the shadow region of a square cylinder, see Fig. 2.24. This theory describes the total field to be the sum of incident, reflected, and diffracted rays. No refracted rays exist since the object is conducting, and therefore cannot be penetrated by the electromagnetic rays.

The field at the receiver resulting from the diffracted ray is found by calculating the incident field at the cylinder corner and then multiplying this by a diffraction coefficient which is a function of the angles (and distances) to transmitter and receiver locations. For example, if the distance from the transmitter (receiver) to the corner is ρ (ρ') with corresponding angle, φ (φ'), then the field at the receiver due to the corner diffraction is given by,

$$E_{rcv} = \frac{-j\omega\mu I}{4}\left[\frac{2}{\pi\gamma\rho}\right]^{1/2} e^{-\gamma\rho} D(\varphi,\varphi',\rho,\rho')\ \frac{e^{-\gamma\rho'}}{\sqrt{\gamma\rho'}} \qquad (2\text{-}63)$$

where γ is the propagation constant of the surrounding medium and D() is the diffraction coefficient function. The total field at the

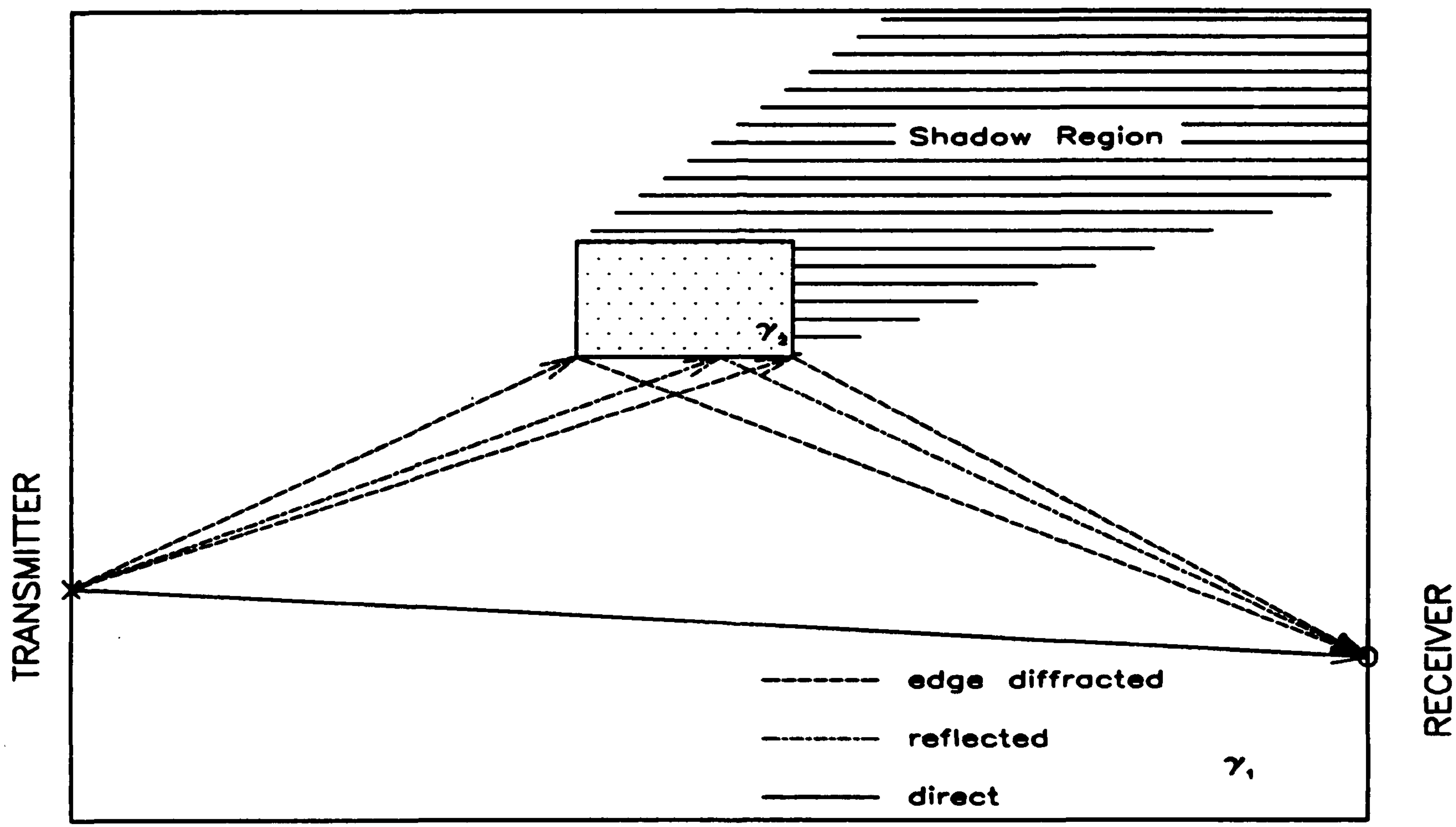


Fig. 2.24. Illustration of diffracting, reflecting, and direct rays for ray optics simulation of scattering from a rectangular cylinder.

receiver is then the sum of the fields from the diffracted rays from all non-shadowed edges of the cylinder, the field from any ray which reflects from the side of the cylinder, and the field from the direct ray from transmitter to receiver. Refer, again to Fig. 2.24.

Fortunately, this theory has been extended to diffractions from wedges (corners) with finite conductivity [42], [43]. The field at the receiver due to the edge diffracted ray is calculated in the same way as for the perfectly conducting object using (2-63), except that the diffraction coefficient is also a function of the surface impedance of the wedge. See [43, eqns. (9) - (15)] for a complete description of the lossy wedge diffraction coefficient.

The surface impedance for the wedge (cylinder, in the present case) is given by,

$$Z_{surf} = \frac{\eta_a}{\eta_e}, \tag{2-64a}$$

$$\eta_a = \left[\frac{j\omega\mu}{\sigma_a + j\omega\epsilon_a}\right], \tag{2-64b}$$

$$\eta_e = \left[\frac{j\omega\mu}{\sigma_e + j\omega\epsilon_e}\right], \tag{2-64c}$$

where η_a and η_e represent, respectively, the intrinsic impedances of the anomaly and the external medium. This surface impedance is the effective impedance seen by the field at the interface to the cylinder [41]. It represents an approximation to the boundary conditions at the interface. This approximation is valid when the magnitude of the refractive index of the cylinder is much greater than the external medium. The complex index of refraction is given by,

$$N = \left[\frac{1}{\epsilon_0}\left\{\epsilon + j\,\frac{\sigma}{\omega}\right\}\right]^{1/2}, \tag{2-65}$$

where ϵ and σ are the electrical parameters of the medium being

considered, while ε_0 is the permittivity of free space. The magnitude of the index of refraction is plotted against frequency, conductivity, and permittivity in Figs. 2.25 - 2.27. As can be seen from these plots, for a background medium having electrical parameters which are typical for dry soil (σ=0.001 S/m, ε_r=10), the anomaly must have a conductivity greater than 0.1 S/m and/or a relative permittivity greater than 50 in order for the index of refraction of the anomaly to be at least twice as large as the index of refraction of the medium. For these values, the surface impedance approximation gives good results when simulating cross-hole data using diffraction theory.

For example, Fig. 2.28 shows the electromagnetic response for a square cylinder in a homogeneous medium. The results were obtained by calculating the electric field at varying depth in a receiver borehole located 20 m from the transmitter borehole. The transmitting antenna is 17 m below the top receiver. The earth between boreholes has electrical parameters of σ=0.001 S/m and ε_r=10. The square cylinder has electrical parameters of σ=0.1 S/m and ε_r=15. It is 2 m on a side, is at a depth of 9 m, and is at a horizontal distance of 11 m from the transmitter borehole. Note how significant the diffracted field is. Also plotted in this figure is the field that was reflected from one face of the cylinder. The value of this reflected field must be found using the reflection coefficient also based on the surface impedance approximation, and given by,

$$\Gamma = \frac{Z_{surf}\cos\theta - 1}{Z_{surf}\cos\theta + 1}, \qquad (2\text{-}66)$$

where θ is the angle of incidence measured from the normal to the interface. This reflection coefficient is analogous to that obtained from transmission line theory assuming that the line impedance is normalized to unity.

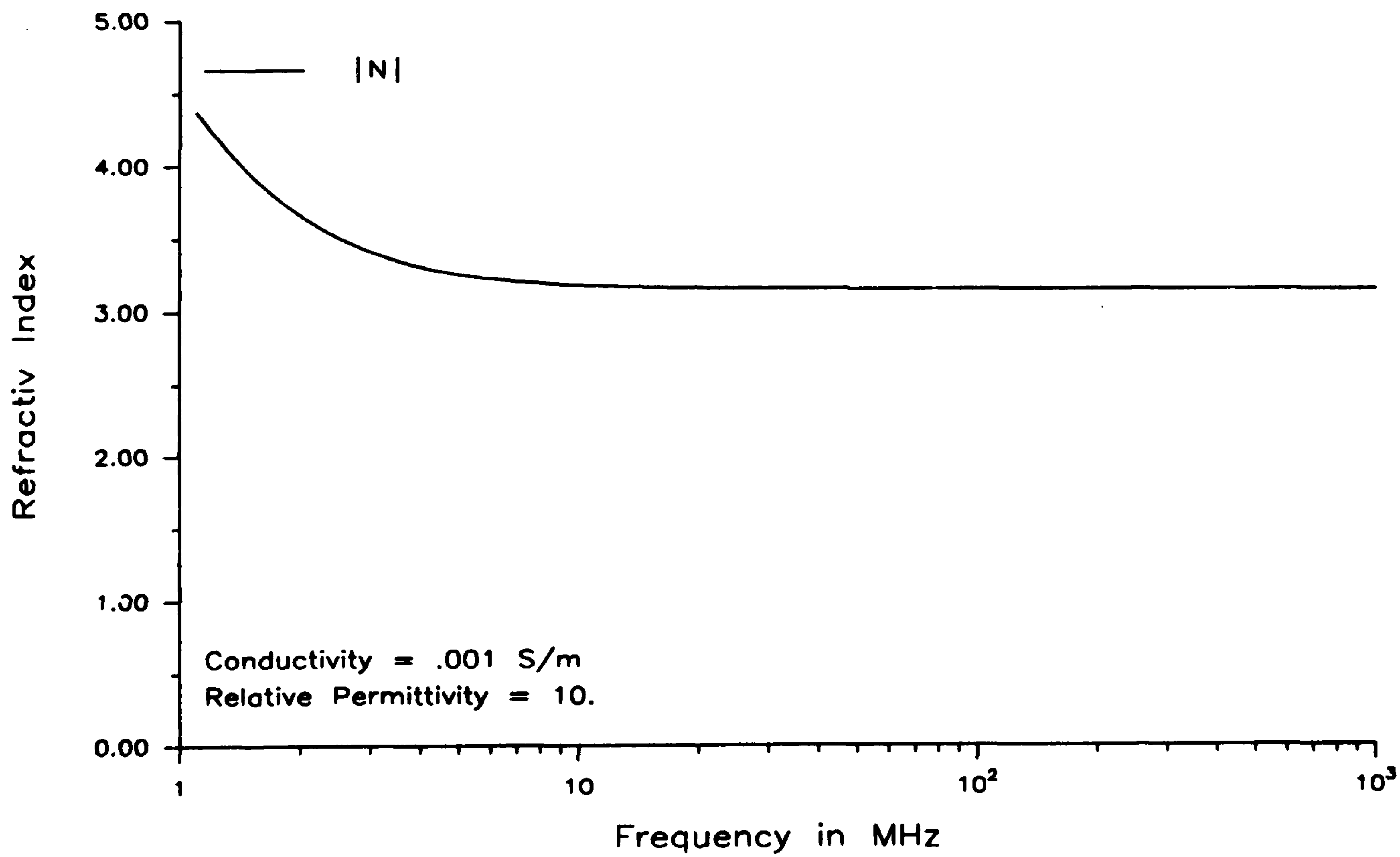


Fig. 2.25. Magnitude of the complex index of refraction plotted against frequency for a medium with conductivity = 0.001 S/m and permittivity = 10.

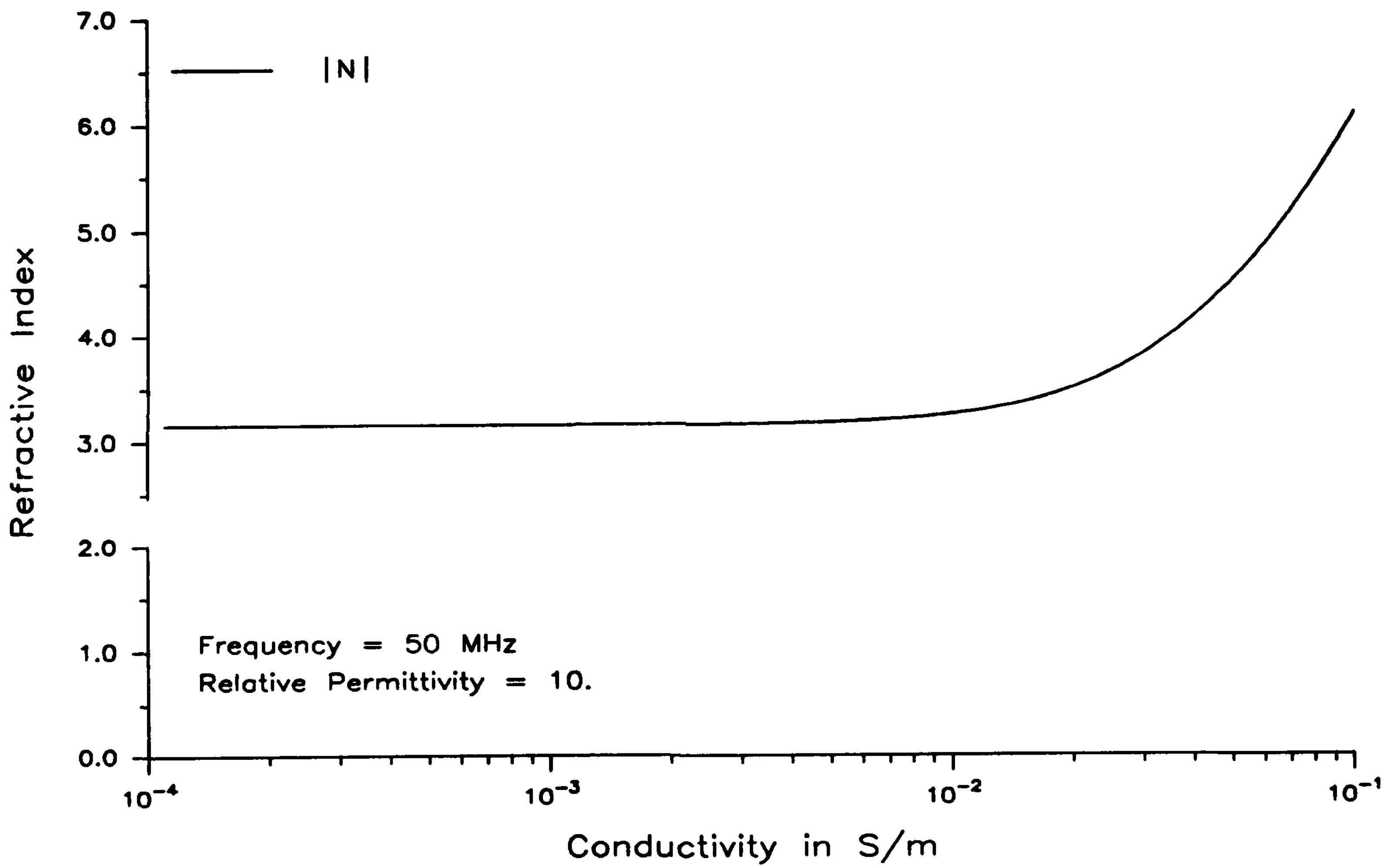


Fig. 2.26. Magnitude of the complex index of refraction plotted against conductivity for a medium with permittivity = 10, and frequency = 50 MHz.

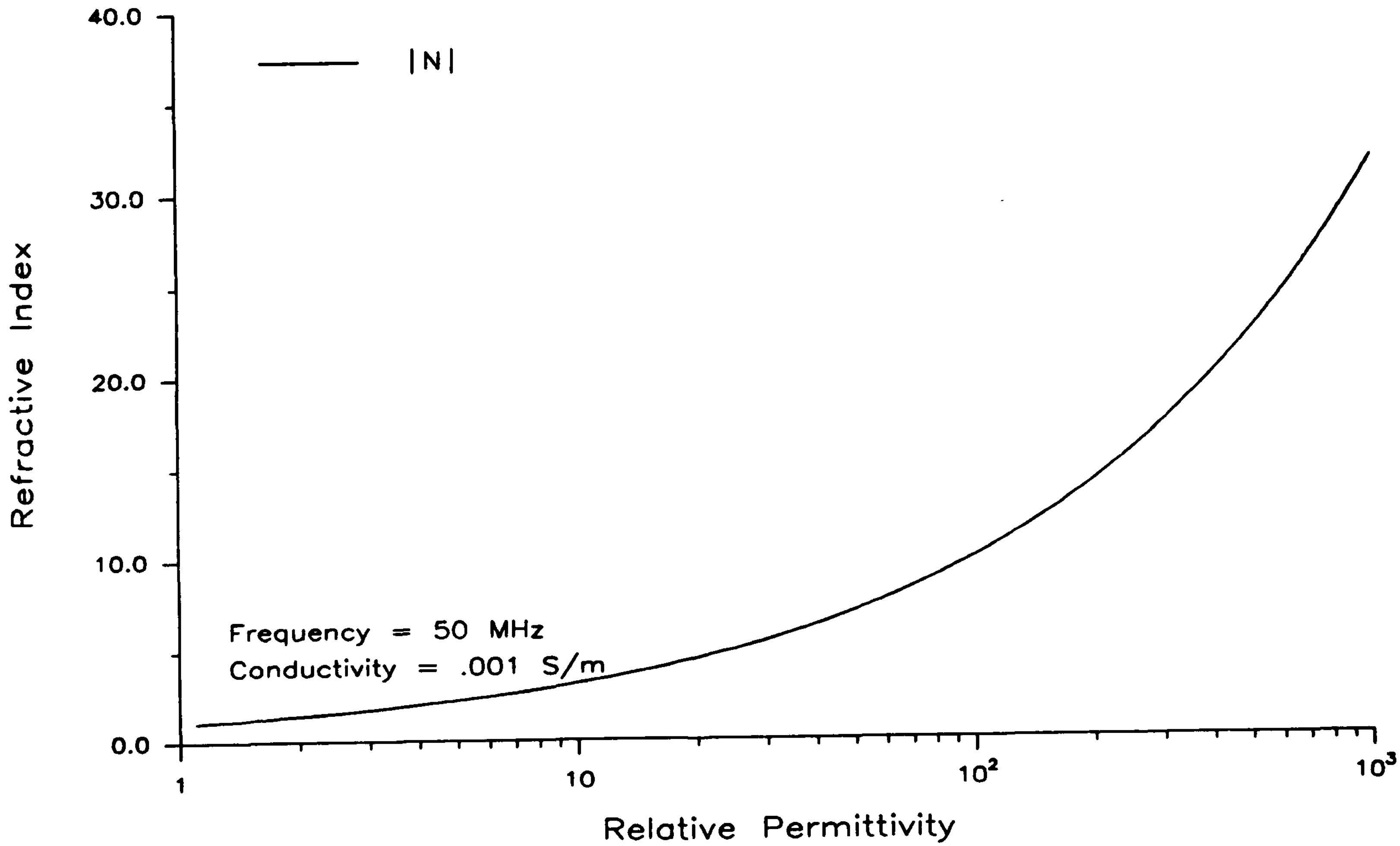


Fig. 2.27. Magnitude of the complex index of refraction plotted against permittivity for a medium with conductivity = 0.001 S/m, and frequency = 50 MHz.

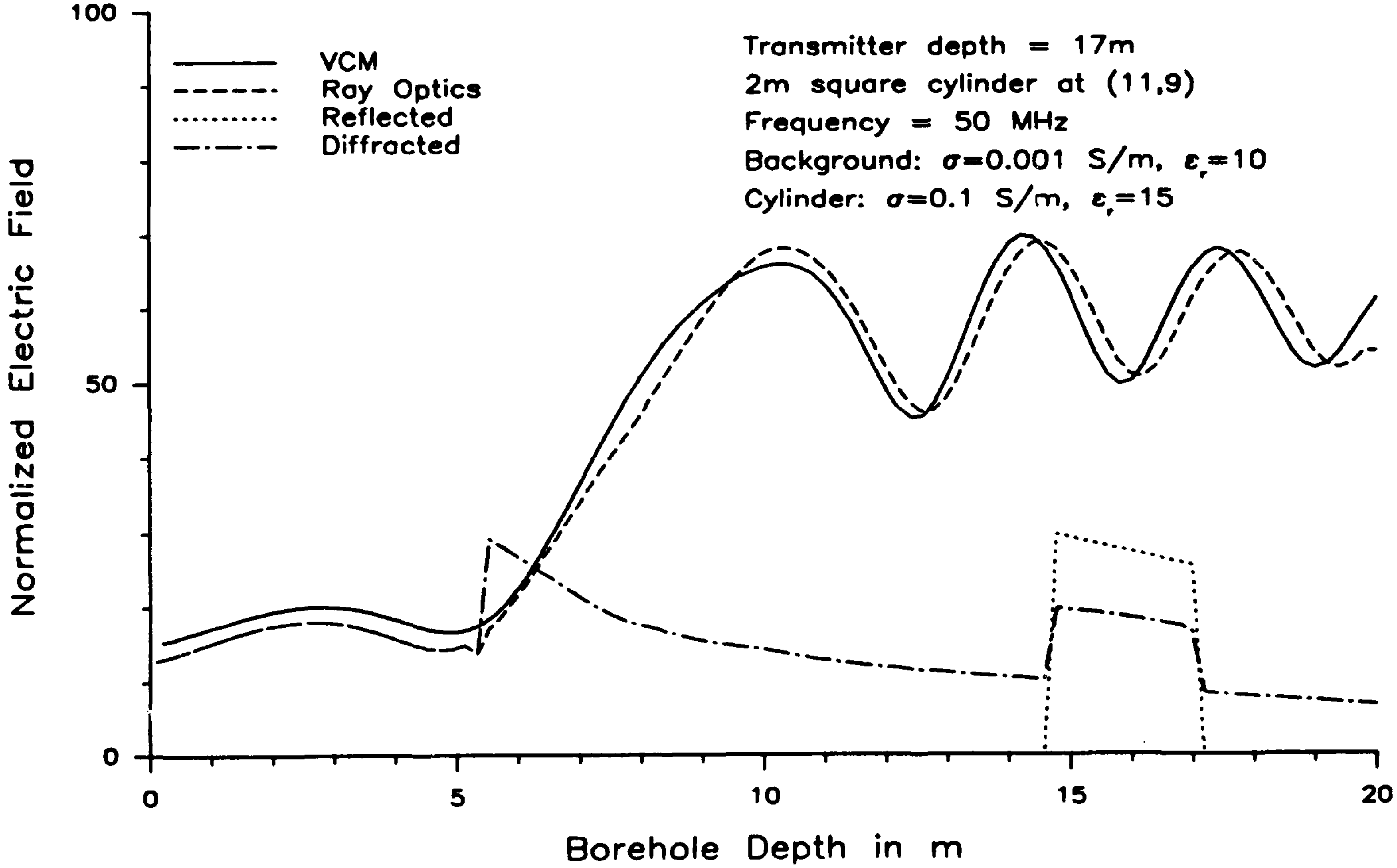


Fig. 2.28. Cross-hole simulations using VCM and ray optics for a square cylinder, in a homogeneous medium.

2.5.4 Summary of Ray Optics Results

It has been shown that ray optics techniques can be used to simulate cross-hole geophysical data. In this simulation process, there are three cases of interest,

a) Case I is a low conductivity cylinder (e.g., a tunnel) imbedded in a homogeneous medium. In this case one must consider direct, reflected, and refracted rays which reach the receiver. The method for finding the total field at the receiver was discussed in Section 2.5.2.

b) Case II is a high conductivity cylinder imbedded in a homogeneous medium. In this case direct, diffracted, and reflected rays must be taken into account. The field at the receiver can be found using the methods presented in Section 2.5.3.

c) Case III is a conducting cylinder in which the index of refraction is not significantly higher than the surrounding medium. Therefore, the impedance boundary approximation cannot be used. However, by studying data from VCM simulations, it is apparent that significant energy is diffracted from the cylinder edges. Therefore, ray optics cannot be used to adequately model this case.

For cases I and II above, ray optics can be used to predict the ray paths the electromagnetic field follows in generating the data. This can be used to determine the reasonableness of assuming the rays follow a straight line path from transmitters to receivers. This assumption is made in the algebraic inversion process described in the next chapter. The straight ray assumption does not take into account reflections, refractions, and diffractions the ray might undergo (of course, this is not a bad assumption given no other *a priori* information).

As an illustration of determining the validity of using the straight ray assumption, refer to Fig. 2.29. This is a plot of the electromagnetic response of a square cylinder in a homogeneous medium. The parameters are the same as for Fig. 2.28, except that the transmitter depth is 6 meters. Along with the curves using VCM and ray optics is the curve obtained using a straight ray assumption. For this last curve, the locations in the receiver borehole in the shadow region of the cylinder indicate small values of the electric field. These small values of the field are the result of assuming ray paths straight through the cylinder (which is highly attenuating). By comparing this 'straight ray' curve to the other curves, the following observations can be made:

a) Since the actual electric field (using VCM or ray optics approximations) is much higher than the 'straight ray' field in the shadow region, when the cross-hole data is inverted, the apparent attenuation of the cylinder will be be much lower than expected.

b) Since the actual electric field in the region adjacent to the shadow does not abruptly return to the incident field value (as does the 'straight ray' field), the apparent size of the cylinder will be larger than expected when the data is inverted.

c) The peaks and nulls in the actual data outside of the shadow region indicate a source of additive noise for the reconstruction process.

In summary, the diffraction and reflection effects will cause errors to be present in the reconstructed image. Some means of reducing these errors will be discussed in the next chapter.

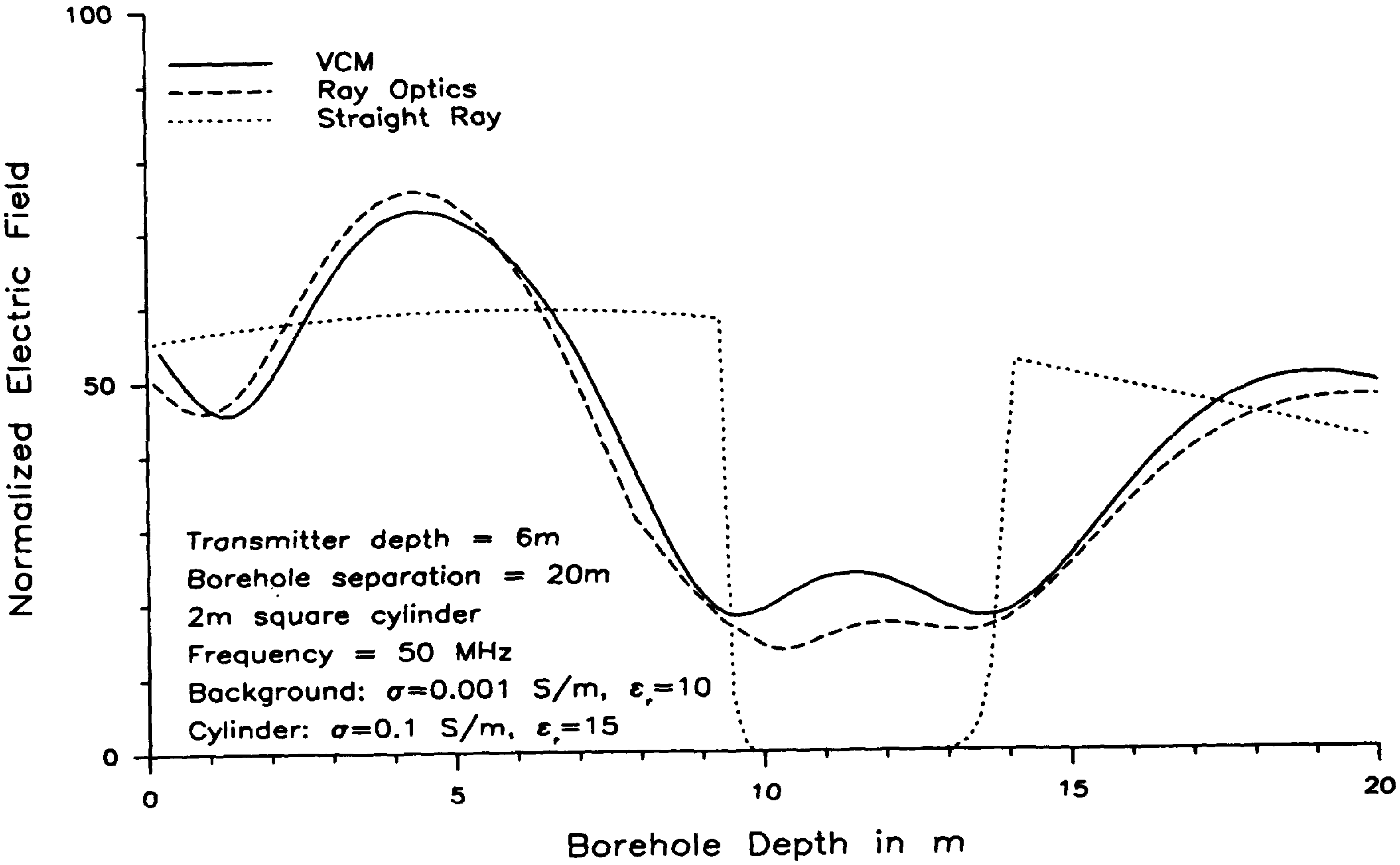


Fig. 2.29. Cross-hole simulations using VCM, ray optics, and straight ray path methods for a square cylinder in a homogeneous medium.

CHAPTER III

IMAGE RECONSTRUCTION METHODS FOR GEOPHYSICAL APPLICATIONS

3.1 Introduction

Now that several models for describing the propagation of electromagnetic waves in the earth have been presented, methods for determining the electrical characteristics of the earth from measurements of received electromagnetic fields can be developed. Once the electrical characteristics of the earth have been found, some interpretations as to the composition of the earth can be made. For example, if a region of high conductivity exists in the earth, then a geophysicist might suspect an oil or mineral deposit in this region. At this point the region would warrant further study.

For the techniques to be developed, it will be assumed that measurements are made using a cross-borehole arrangement (although the techniques will be applicable in situations when either transmitters or receivers are mounted on the earth's surface). For an illustration of the measurement process, refer again to Fig. 1.1. If the assumption is made that the electromagnetic waves travel in the plane of the two boreholes, then the problem is a two-dimensional one. In this case, one can speak of the cross sectional image of the earth between the boreholes. This image is, in fact, just some electrical parameter of the earth (e.g., attenuation or index of refraction) as a function of position (i.e., x and y coordinates). Although it may seem overly restrictive to assume a two-dimensional model, this is an adequate assumption in many cases. For example, if the goal of a geophysicist is to locate a tunnel in a region between two boreholes, then the axial direction of the tunnel is probably known, and one would need only determine its depth and horizontal distance from either borehole. In this case, the reconstructed image would ideally show a cross section

of the tunnel located in the earth. Usually some kind of gray scale is used to generate the image (with, for example, different levels representing different attenuation values), although contour and three-dimensional plots are also used.

Now that the image reconstruction problem has been described in this geophysical setting, the objective is to solve the problem using some of the models developed in the Chapter II. In particular, the Born approximation and the ray optics models will be used.

3.2 A Comparison of Reconstruction Methods

3.2.1 Straight Ray Approximations

The straight ray model is the crudest model used in Chapter II, and accordingly, it can be used to generate the simplest reconstruction algorithms. The assumption is made that the electromagnetic waves travel along straight ray paths connecting transmitters and receivers. Reflection, refraction, and diffraction effects are ignored. Reconstruction techniques using the straight ray assumption are usually based on finding a relationship between the line integral of the parameter of interest and the measurement data. This relationship can be obtained for continuous wave (CW) measurements in the following manner (an analogous development for time-of-flight measurements can also be made).

a) For the line source antenna, the electric field at a radial distance ρ from the antenna is given by equation (2-53), which is repeated here in a simplified form,

$$\begin{aligned} E &\sim K \frac{e^{-\gamma\rho}}{\sqrt{\rho}}, \\ &= \frac{K}{\sqrt{\rho}} e^{-\alpha\rho} e^{-j\beta\rho}, \end{aligned} \qquad (3\text{-}1)$$

where the far field approximation has been made, and K is a complex scale factor whose exact form is given in (2-53).

b) In CW tomography, measurements are made of the received electromagnetic power; therefore, only the magnitude of the field is of interest. The magnitude of the field is given by

$$|E| \sim |K| \frac{e^{-\alpha\rho}}{\sqrt{\rho}}. \tag{3-2}$$

c) If the assumption is made that α is constant for small changes in ρ, then the differential change in the electric field is given by,

$$\Delta|E| = -\,|K|\left[\alpha \frac{e^{-\alpha\rho}}{\sqrt{\rho}} + \frac{e^{-\alpha\rho}}{2\rho^{3/2}}\right]\Delta\rho. \tag{3-3}$$

from which we obtain

$$\frac{\Delta|E|}{|E|} = -\left[\alpha + \frac{1}{2\rho}\right]\Delta\rho. \tag{3-4}$$

d) The above equation can now be used to find a line integral relation between the attenuation and the electric field as,

$$\int_{E_0}^{E_1} \frac{d|E|}{|E|} = \int_{r_0}^{r_1} -\alpha(\rho)\, d\rho - \frac{1}{2}\int_{r_0}^{r_1} \frac{d\rho}{\rho}, \tag{3-5}$$

or

$$\ln(E_1\sqrt{r_1}) - \ln(E_0\sqrt{r_0}) = \int_{r_0}^{r_1} -\alpha(\rho)\, d\rho, \tag{3-6}$$

where the line integral is over a particular radial ray path

and E_1 (E_0) is the magnitude of the electric field at a radial distance r_1 (r_0).

For the development above, the receiver is assumed to be at a radial distance r_1. The radial distance r_0 must be sufficiently large for the far field approximation to be valid. If the term $E_0\sqrt{r_0}$ is normalized to unity (by adjusting the gain of the transmitting antenna), then the equation above reduces to

$$\ln(E_1\sqrt{r_1}) = \int_{r_0}^{r_1} -\alpha(\rho)\, d\rho. \tag{3-7}$$

The reconstruction problem is seen to consist of finding the attenuation as a function of position, from knowledge of the electric field intensity E_1. It is easy to see that $\alpha(\rho)$ will not be uniquely determined by measuring E_1. In order to resolve this uniqueness problem, measurements must be taken over a large number of ray paths. Fig. 3.1 illustrates a set of measurements taken along ray paths which are at an angle of $(\pi/2 - \phi)$ with respect to the x axis. One such ray path is labeled L in the figure. This ray path is determined by the angle ϕ, and by its radial distance, p, from the origin. For this ray path, (3-7) can be rewritten as,

$$\ln(E_L\sqrt{r_1}) = \int_{L(p,\phi)} -\alpha \; d\ell, \tag{3-8}$$

or

$$\ln(E_L\sqrt{r_1}) = - \mathcal{R}_\phi(\alpha), \tag{3-9}$$

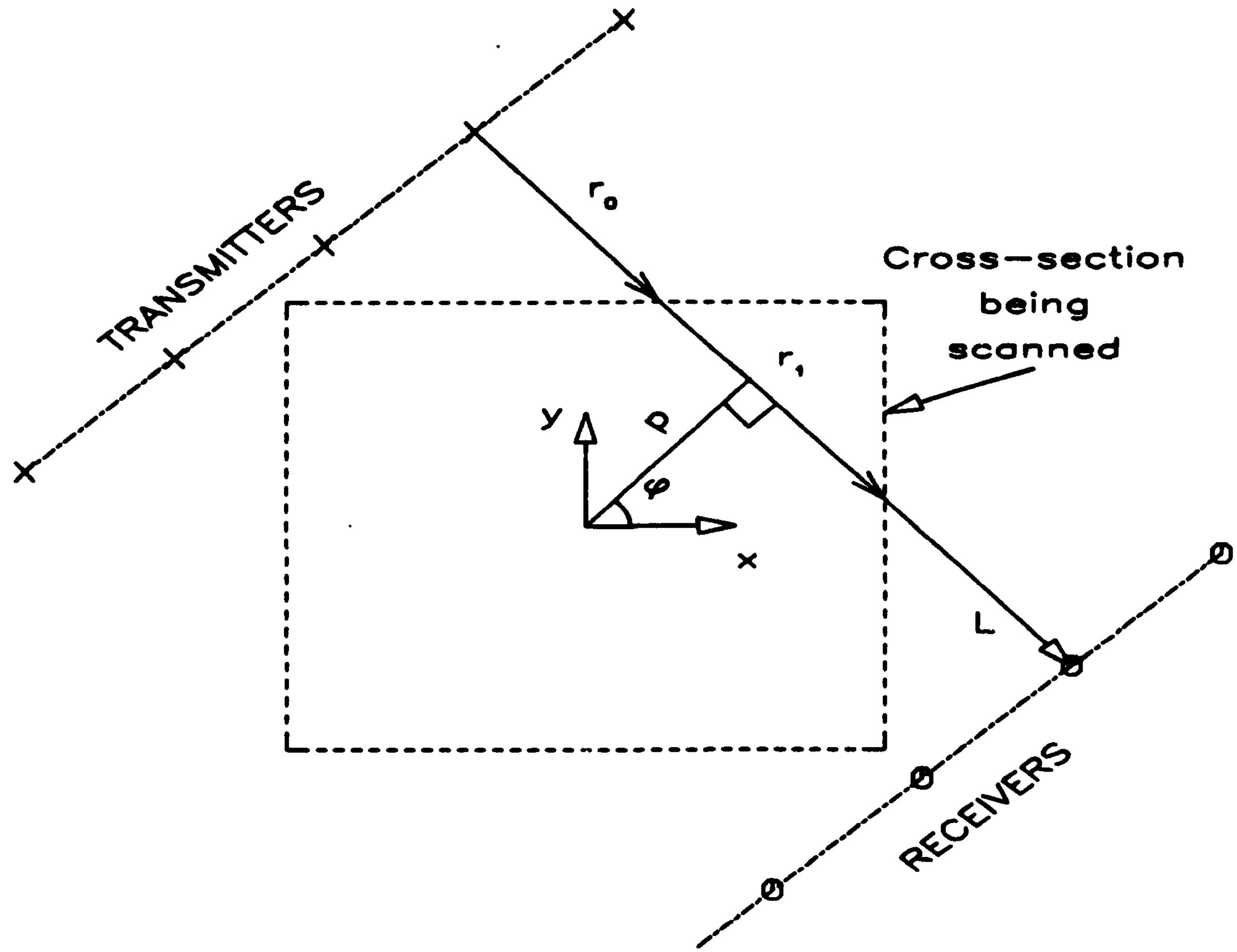


Fig. 3.1. Illustration for obtaining a set of projection measurement data for a fixed angle.

where E_L indicates the magnitude of the electric field at the receiver end of the line L, $d\ell$ is the incremental length along the line, and $\mathcal{R}_\phi$ is called the Radon transform operator along the line L at the angle ϕ.

If a set of measurements is taken for all p such that $L(p,\phi)$ intersects the region, then this set is called a projection of the region at an angle ϕ. If projections are taken for all ϕ in the range $[0,\pi]$, then $\alpha(x,y)$ will be uniquely determined in the region [44], and a simple and efficient algorithm exists (i.e., the convolution-backprojection algorithm) for recovering the attenuation profile.

Unfortunately, in the present setting we are faced with an incomplete data problem in that due to physical constraints, measurements can not be made for all values of p or ϕ. In particular, Fig. 3.1 would not be accurate in that projections can only be taken at values of ϕ close to 90 degrees (assuming that the boreholes are located on the sides of the region). In this case, the convolution-backprojection algorithm is inadequate for reconstructing the image of attenuation values. In such limited data cases it is generally accepted that the inversion problem is highly ill-posed, and therefore some means of regularization or incorporation of *a priori* information must be used in order to obtain meaningful results [45]. Methods for better posing this inversion problem are discussed in a later section, while some of the more well-known inversion techniques based on (3-8) are discussed below.

A. *Radon Transform Theory and the Convolution-Backprojection Algorithm*

The Radon transform theory developed here is taken from Deans [44]. The description below will be brief since the inversion process obtained is not useful for the limited data reconstruction problem. The primary interest here is to show that a direct solution of the reconstruction problem exists for the full data case. As

mentioned, the Radon transform of a function in the x-y plane is the set of integrals along all lines passing through the support of this function. That is,

$$\mathcal{R}[\alpha(x,y)] = \int_L \alpha(x,y)d\ell \tag{3-10}$$

for all lines L. This defines the Radon transform. We will make the convention to call (3-10) the Radon operator when the restriction of 'for all lines L' is not imposed, although this distinction may become fuzzy at times. In order to express (3-10) in a more usable form, we define the line L in the normal form (Fig. 3.1),

$$p = x\cos\phi + y\sin\phi \tag{3-11}$$

where p is the normal distance from the origin to L, and ϕ is the angle between the normal and the x-axis. We can use the Dirac delta function to allow the integration to be performed over this line as

$$\begin{aligned}\mathcal{R}[\alpha(x,y)] &= \int_a^b\int_a^b \alpha(x,y)\delta(p-x\cos\phi-y\sin\phi)dxdy \\ &= \int_a^b\int_a^b \alpha(x,y)\delta(p-\langle\mathbf{X},\boldsymbol{\xi}\rangle)dxdy,\end{aligned} \tag{3-12}$$

where $\mathbf{X} = (x\ y)^T$, $\boldsymbol{\xi} = (\cos\phi\ \sin\phi)^T$, and it is assumed that the attenuation is measured on the cross sectional region $[a,b]\times[a,b]$. We are taking some liberties in this development, such as using the Dirac delta 'function', but this derivation can be made rigorous by appealing to the theory of distributions (see, for example, [46]). The Radon transform is seen to be a function of p, parameterized by the angle ϕ. It may be useful, though, to consider the transform as a function of both p and ϕ:

$$\mathcal{R}[\alpha(x,y)] = \check{\alpha}_{\phi}(p) = \check{\alpha}(p,\phi), \tag{3-13}$$

where '$\check{\ }$' denotes the Radon transform.

Here it is assumed that the Radon transform is a mapping from the set of square integrable functions in the x-y plane to the set of functions defined on the semi-infinite cylinder ($\mathbb{R}\times S^1$). To make this clear, let p=0, then the cylinder reduces to a circle and all integrals are over lines through the origin in the x-y plane. The Radon transform in this case will be the values of the line integrals at all points ϕ around the circle.

The ultimate goal in developing Radon transform theory is to be able to determine the function α in the x-y plane given its projections which are usually discrete samples of the Radon transform of α. That is, for the case at hand, we want to be able to determine the attenuation constant function over the cross section of the earth given a set of received electromagnetic power data points. Therefore, the existence of the inverse Radon transform needs to be determined, along with a method for calculating the inverse.

The development of the inverse Radon transform presented here is based on the relationship between the Radon and Fourier transforms. This relationship is known as the projection-slice theorem [44]. The theorem is easily obtained by taking the Fourier transform of the function $g(p,\phi)$ (which is assumed to be the Radon transform of some attenuation function) with respect to the p variable as

$$\begin{aligned} g^{\wedge}(p,\phi) &= [\mathcal{R}(\alpha)]^{\wedge}(p,\phi) \\ &= \int_{-\infty}^{\infty}\int_a^b\int_a^b \alpha(x,y)\delta(p-\langle\mathbf{x},\xi\rangle)dxdye^{j\omega p}dp \\ &= \int_a^b\int_a^b \alpha(x,y)\int_{-\infty}^{\infty}\delta(p-\langle\mathbf{x},\xi\rangle)e^{j\omega p}dpdxdy \end{aligned}$$

$$= \int_a^b\int_a^b \alpha(x,y)e^{j\omega<\mathbf{x},\xi>}dxdy \qquad (3\text{-}14)$$

where '^' signifies the Fourier transform operation. Switching the order of integration is justified since α is assumed to be a square integrable function of compact support. In words this theorem states that a Radon transform followed by a Fourier transform of the function α is equivalent to taking the two-dimensional Fourier transform of α.

This theorem leads to two methods of inverting the Radon transform. The first follows immediately from (3-14), that is, take the Fourier transform of the Radon transform (of α), and perform a two-dimensional Fourier inversion of the result, thereby recovering α. This results in

$$\alpha(p,\phi) \quad = \quad \mathbf{F}_2^{-1}[\mathbf{F}(\mathcal{R}\{\alpha(x,y)\})] \qquad (3\text{-}15)$$

where α is found as a function of polar coordinates. In (3-15), $\mathbf{F}$ stands for the Fourier transform operator and $\mathbf{F}_2^{-1}$ stands for a two-dimensional Fourier inversion. This method has the drawback that the attenuation is determined on a polar grid, and therefore must be interpolated to find α as a function of (x,y).

The second method is the one given by Radon [5], but the development here is patterned after Rowland [47]:

$$\begin{aligned}
\alpha(x,y) &= \mathbf{F}_2^{-1}\left\{\mathbf{F}_2(\alpha)\right\} \\
&= \int_0^{2\pi}\int_0^{\infty}\mathbf{F}_2(\alpha)(\rho\cos\theta,\rho\sin\theta) \\
&\qquad \exp[2\pi j\rho(x\cos\theta+y\sin\theta)]\rho d\rho d\theta \\
&= \int_0^{\pi}\int_{-\infty}^{\infty}\mathbf{F}_2(\alpha)(\rho\cos\theta,\rho\sin\theta) \\
&\qquad \exp[2\pi j\rho(x\cos\theta+y\sin\theta)]|\rho|\rho d\rho d\theta
\end{aligned}$$

$$= \int_0^{\pi}\int_{-\infty}^{\infty} \mathbf{F}(\mathcal{R}(\alpha))(\rho,\theta)\,|\rho| \exp[2\pi j\rho(x\cos\theta+y\sin\theta)]\rho d\rho d\theta$$

$$= \int_0^{\pi}\int_{-\infty}^{\infty} [\,|x|\ \mathbf{F}(\mathcal{R}(\alpha))](\rho,\theta) \exp[2\pi j\rho(x\cos\theta+y\sin\theta)]\rho d\rho d\theta$$

$$= \int_0^{\pi} \mathbf{F}^{-1}[\,|x|\ \mathbf{F}(\mathcal{R}(\alpha))](x\cos\theta+y\sin\theta,\theta)d\theta. \qquad (3\text{-}16)$$

If a back-projection operator is defined by

$$[\mathcal{B}h](x,y) = \int_0^{\pi} h(x\cos\theta+y\sin\theta,\theta)d\theta \qquad (3\text{-}17)$$

then (3-16) can be written as

$$\alpha(x,y) = \mathcal{B}\mathbf{F}^{-1}[\,|x|\ \mathbf{F}(\mathcal{R}(\alpha))] \qquad (3\text{-}18)$$

From the convolution theorem for Fourier transforms

$$-2\pi\mathbf{F}^{-1}[\,|x|\ \mathbf{F}(\mathcal{R}(f))] = \mathcal{H}\ \mathcal{D}\ \mathcal{R}(\alpha) \qquad (3\text{-}19)$$

where $\mathcal{H}$ is the Hilbert transform operator and $\mathcal{D}$ is the differential operator. Then (3-18) becomes

$$\alpha(x,y) = \frac{-1}{2\pi}\ \mathcal{B}\ \mathcal{H}\ \mathcal{D}\ \mathcal{R}(\alpha) \qquad (3\text{-}20)$$

which is equivalent to the inversion formula developed by Radon [5]. The convolution-backprojection method uses (3-20) as the basis for inverting the sampled Radon transform. In this method a convolution is used instead of the Hilbert transform and differentiation in (3-20)

(see [48] for details).

B. The Algebraic Method for Solving the Image Reconstruction Problem

In this section a simpler method is presented for solving the reconstruction problem. Note that in deriving this method, some assumptions must be made which might not always be valid. This method is easily developed by first discretizing the cross sectional region being scanned [48]. Fig. 3.2 shows one such discretization obtained by dividing the region into rectangular picture elements (pixels). If the assumption is made that the attenuation is constant over each pixel, then one can speak of an image vector of attenuation values as

$$x = [\alpha_1 \;\; \alpha_2 \;\; \cdots \;\; \alpha_n]^T, \tag{3-21}$$

where α_i is the attenuation value of the i^{th} pixel and 'T' denotes matrix transpose. For the example in Fig. 3.2, the pixels are numbered consecutively by rows, and n=16. For this discretization, the line integral in (3-8) reduces to an algebraic equation. For the line integral over the line L in Fig. 3.2, this algebraic equation is

$$\int_L \alpha \, d\ell = d_3\alpha_3 + d_4\alpha_4 + d_8\alpha_8, \tag{3-22}$$

where the d's represent the ray path distances through the respective pixels. If d_{ij} is now defined by the distance the i^{th} ray path travels through the j^{th} pixel, then the following equation is obtained

$$\beta_i = \sum_{j=1}^{n} d_{ij} \, \alpha_j, \tag{3-23}$$

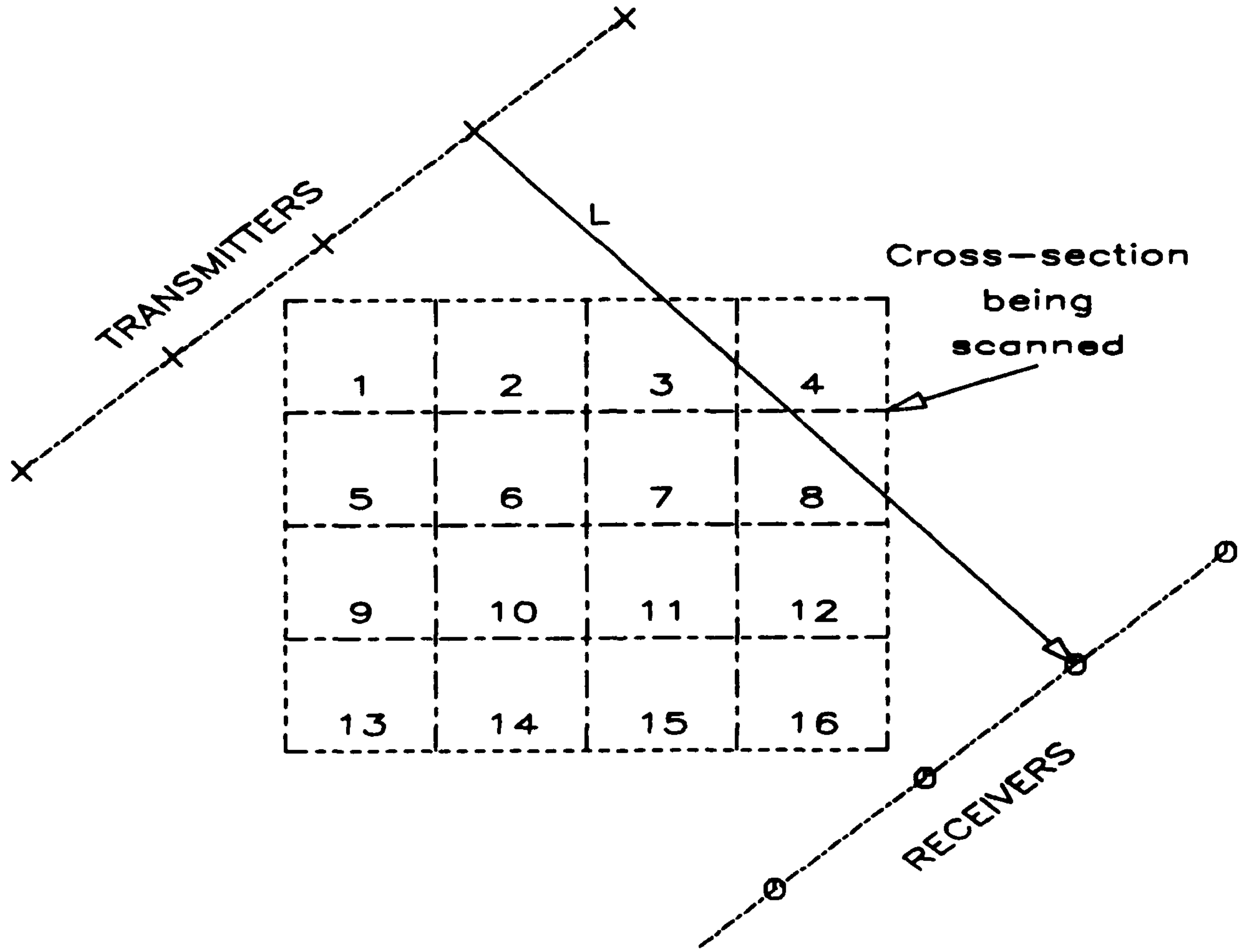


Fig. 3.2. Illustration of the discretization process for deriving the algebraic inversion method.

where β_i is the measurement data for the i^{th} ray path, and is equal to the left side of (3-8). Note that d_{ij} will be zero if the i^{th} ray path does not intersect the j^{th} cell.

We are now in a position to develop a matrix equation relating the measurement data to the unknown attenuation values by defining

$$b = [\beta_1 \quad \beta_2 \quad \cdots \quad \beta_m]^T, \tag{3-24}$$

as the measurement vector, and

$$A = [d_{ij}], \tag{3-25}$$

as the path length matrix. The matrix equation is

$$b = Ax + e, \tag{3-26}$$

where an error vector, e, has been added to account for measurement and modeling errors. The basis for the algebraic method for image reconstructions is (3-26). Algorithms which solve for the unknown image vector, x, will be presented in Chapter IV.

3.2.2 Diffraction Tomography

Diffraction tomography was developed as an improvement over straight-ray methods in that ray diffractions (as well as reflections and refractions) are implicitly included in the inversion process. As mentioned in the Chapter II, the forward modeling method using the Born approximation (the Rytov approximation can also be used) is the basis for the inversion algorithm known as diffraction tomography [19], [37]. The integral formulation for the scattered field was given in (2-48).

This equation is repeated here in simplified form

$$E^s(\mathbf{x}) \sim \iint_A O(\mathbf{x}')E^i(\mathbf{x}')G(|\mathbf{x}-\mathbf{x}'|)d\mathbf{x}', \tag{3-27}$$

where $O(\mathbf{x})$ represents the object inhomogeneity. The Green's function in this equation is in fact the zero order Hankel function of the second kind. It has an integral expansion given by [37]

$$\begin{aligned} G(|\mathbf{x}-\mathbf{x}'|) &= \frac{j}{4} H_0^{(2)}(-j\gamma_e|\mathbf{x}-\mathbf{x}'|) \\ &= \frac{j}{4\pi}\int_{-\infty}^{\infty} \frac{1}{\eta}\, e^{-j[\lambda(x-x')+\eta(y-y')]}d\lambda, \end{aligned} \tag{3-28}$$

where

$$\eta := j\,(\gamma_e^2 + \lambda^2)^{1/2}. \tag{3-29}$$

As in Chapter II, γ_e is the propagation constant for the external medium. If this expansion is substituted into (3-27) we get

$$E^s(\mathbf{x}) \sim \frac{j}{4\pi}\iint_A O(\mathbf{x}')E^i(\mathbf{x}')\int_{-\infty}^{\infty} \frac{1}{\eta}\, e^{-j[\lambda(x-x')+\eta(y-y')]}d\lambda d\mathbf{x}'. \tag{3-30}$$

In order to simplify the derivation, it is assumed that we have a plane wave incident on the object of the form

$$E^i(\mathbf{x}) = e^{-\gamma_e y}, \tag{3-31}$$

where, for convenience, the plane wave is taken to be traveling in the

+y direction. For an extension of these results to the case of a cylindrical input wave front, see [37]. Under the assumption that the scattered field is measured along the line $y=\ell_0$, (3-30) becomes

$$E^s(\mathbf{x}) \sim \frac{j}{4\pi} \iint_A \int_{-\infty}^{\infty} \frac{O(\mathbf{x}')}{\eta} e^{-\gamma_e y'} e^{-j[\lambda(x-x')+\eta(\ell_0-y')]} d\lambda d\mathbf{x}' . \tag{3-32}$$

This equation can be Fourier transformed to yield (see [37] for details)

$$\hat{E}^s(\lambda, l_0) = \begin{cases} \frac{j}{2\eta} e^{-j\eta l_0} \hat{O}(\lambda, \eta-\gamma_e), & \text{for } |\lambda| < \gamma_e \\ 0, & \text{for } |\lambda| > \gamma_e \end{cases} \tag{3-33}$$

where η is given in (3-29). This result is in fact a generalization projection-slice theorem in which the Fourier transform of the received field is related to the Fourier transform of the object. In this case, however, the Fourier transform of the object is taken over circular arcs in the frequency domain (see [37] for an illustration). One drawback to the diffraction tomography method is that it is limited to objects which satisfy either the Born or Rytov approximations. This limitation will be discussed later in this chapter.

3.2.3 Other Reconstruction Methods

The methods presented in this section are not as practical as those described in the previous sections. Their limitations are mainly due to excessive computational requirements in terms of time and memory size. Therefore, although these methods may not be practical at the

present time, it is worthwhile giving brief descriptions in anticipation of evolving computer systems.

A. Moment Method Inversions

These methods have been developed for medical applications, where it is desirable to use low-level electromagnetic waves as an alternative to X-rays [49], [50]. The basis for moment method inversion is the VCM (a type of moment method described in Chapter II) used for solving the problem of scattering from an arbitrarily shaped cylinder. The problem was solved by considering the total field to be the sum of an incident field (in the absence of the cylinder), and a scattered field. The unknown scattered field was then represented as a series of pulse basis functions over the area of the cylinder. This scattered field was then substituted into an integral equation of the form of (2-46). After some manipulations, the scattered field is obtained via a matrix equation relating the scattered field, incident field, and the cylinder geometry.

For the inverse problem, the assumption is made that the magnitude and phase of the scattered field are known from measurements, and the goal is to find the location and characteristics of the scatterers. The scatterers are the small circular cylinders into which the region is divided (refer, again, to Fig. 2.16). The inversion method consists of (pseudo-) inverting the matrix equation to find the scatterers. This method has the following difficulties:

1) For a large cross sectional area, the matrix involved in the solution will be very large, resulting in numerical problems. Additionally, the authors in [49] noted that the matrix becomes more ill-conditioned as its size increases.
2) In [49] and [50] the authors assumed scattering of a plane wave. The cross-borehole geometry being considered here involves the scattering of cylindrical waves originating from

multiple sources (transmitting antennas). The challenge would be to use the solution from each transmitter as some kind of *a priori* information to aid in solving the inversion of successive transmitters.

3) Inherent in this inversion process is the need to measure both magnitude and phase of the received signal. However, phase measurements are generally more difficult to obtain than magnitude measurements.

In spite of these difficulties, the moment method inversion technique has the advantage of accounting for diffraction (as well as reflection and refraction) effects without being limited by the restrictions of the Born or Rytov approximation. Therefore, it is an attractive method which might be useful in the future.

B. Model Matching Using the Ray Optics Method

The inversion algorithm being proposed here is shown in block diagram form in Fig. 3.3. The basic idea is to attempt to match the output of a ray optics simulation program to actual field measurement data. This method is similar to that discussed in [51], where the authors were concerned with inversion of seismic data. The differences between their approach and the one given in this section are summarized below.

1) In Fig. 1 of [51], the authors consider matching the model to the data through human intervention. Here, it is suggested to perform this model matching process using a computer.
2) In discussing forward modeling techniques, the authors in [51] mention ray optics methods, but do not include diffraction effects. These effects will have major impact in many cases, and therefore are not ignored here.

The major feature of the method being presented is that it is based on ray optics techniques, which when diffraction effects are

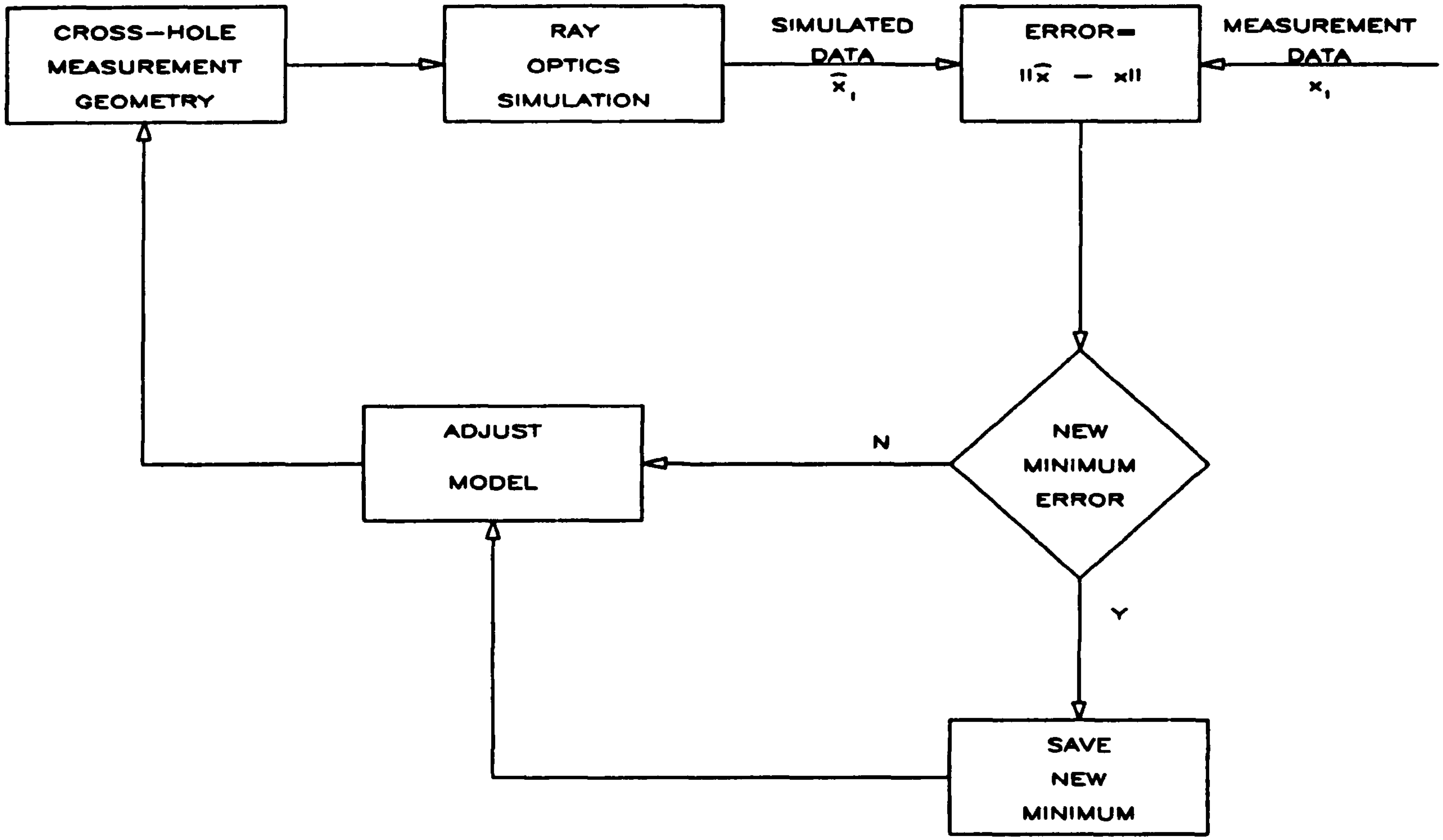


Fig. 3.3. Block diagram for ray optics model matching procedure for geo-inversions.

included, afford an accurate and efficient forward modeling algorithm. Note that although a forward modeling algorithm such as the VCM could be used instead of ray optics in this procedure, it is so computationally intensive that it would make the method infeasible.

The characteristics of this inversion method are listed in the following.

1) The main assumption is that the region under study is homogeneous except for a single scatterer (anomaly) and that the electrical characteristics of the background and scatterer are known.
2) The model is adapted by placing a grid over the cross sectional region, and then locating the scatterer at successive locations on the grid.
3) For each location of the scatterer, the simulated measurement data is compared against the actual measurement data, and the location of the scatterer giving the smallest error is saved.
4) After the smallest error location is found, the process can be repeated over a more finely gridded area centered on this location. In addition, the size and shape of the scatterer can also be adjusted to minimize the error.

The assumption in 1) may seem overly restrictive, but it is possible that a geophycisist has some reason to suspect a particular kind of anomaly (e.g. tunnel or ore deposit) in a homogeneous medium, and would therefore know beforehand its characteristics. In the case of a tunnel, he may even know its approximate size and shape. Additionally, this method could be used in conjunction with some other inversion algorithm in order to refine estimates of an anomaly's location, size, and shape (this procedure will be discussed in more detail in Chapter V). Also, the assumption of a single scatterer could be relaxed by using the method iteratively to find all scatterers. Of course, this would complicate the modeling since the interactions between scatterers would have to be taken into account.

In order to justify modifying the model by moving the cylinder on a grid, one would have to show that the measured field data uniquely determines the location of the cylinder. In addition, one would have to show that for small changes in the cylinder location, the measured data experiences a small change in some norm. Certainly, if the location, size, and shape of the cylinder were all free to change, then the amplitude measurement data would not uniquely determine the cylinder for a single transmitting antenna. Whether this remains true for measurement data from multiple transmitter locations is still a topic of open debate [52]. As for the question of the effect of small changes in the cylinder location, we do know that the diffraction coefficient [43] is a continuous function of angle and distance. However, it would be very difficult to assess the effect a change in location has on the total field (which is the sum of direct, diffracted, and reflected rays). Some indication that a small change in location means a small difference in the total field can be given by considering the scattering from a circular cylinder, which was presented in Chapter II.

The scattered field from a cylinder was given by (2-37), and is repeated here for convenience

$$E^{s}(\rho) = \sum_{n=-\infty}^{\infty} d_n H_n^{(2)}(-j\gamma_e\rho) H_n^{(2)}(-j\gamma_e\rho_t), \tag{3-34}$$

where the origin is taken at the cylinder axis. Chapter II notes that for the cases we are considering, only a finite number of terms are needed in the summation. To simplify the analysis, we assume that the transmitter, cylinder axis, and receiver are colinear, and that the cylinder is moved a distance $\Delta\rho$ along this line. The Hankel functions in (3-34) are analytic off the negative real axis, and, subsequently, a finite sum of these functions will also be analytic in this region. Therefore, (3-34) can be expressed as

$$E^S(\rho) \simeq f(\rho, \rho_t), \tag{3-35}$$

where f is an analytic function equal to the finite sum. From this one can obtain

$$\begin{aligned} E^S(\rho+\Delta\rho) - E^S(\rho) &= f(\rho+\Delta\rho, \rho_t-\Delta\rho) - f(\rho, \rho_t) \\ &= \Delta\rho^2 \, \frac{f(\rho+\Delta\rho, \rho_t-\Delta\rho) - f(\rho, \rho_t)}{\Delta\rho^2} \\ &\sim \Delta\rho^2 \, \frac{\partial^2 f}{\partial\rho\partial\rho_t}, \end{aligned} \tag{3-36}$$

which, since f is analytic, shows that small changes in the location imply small changes in the received field. Of course, to make this analysis complete, the above calculations would have to be performed for arbitrary transmitter and receiver locations. In any case, (3-36) gives some confidence in the method. In support of this, Fig. 3.4 shows the effect on the total field of moving a square cylinder 1 m horizontally in both directions. The results add credence to the claim, since only slight changes in the field are observed.

In testing for a minimum error we choose the norm for square integrable functions given by

$$(\|f-g\|_2)^2 = \int |f(y) - g(y)|^2 dy. \tag{3-37}$$

Since the received field will be measured at discrete points in the receiver borehole, this norm is approximated by

$$(\|f-g\|_2)^2 = \sum_{n=1}^{N} |f(y_i)\ ) - g(y_i)|^2. \tag{3-38}$$

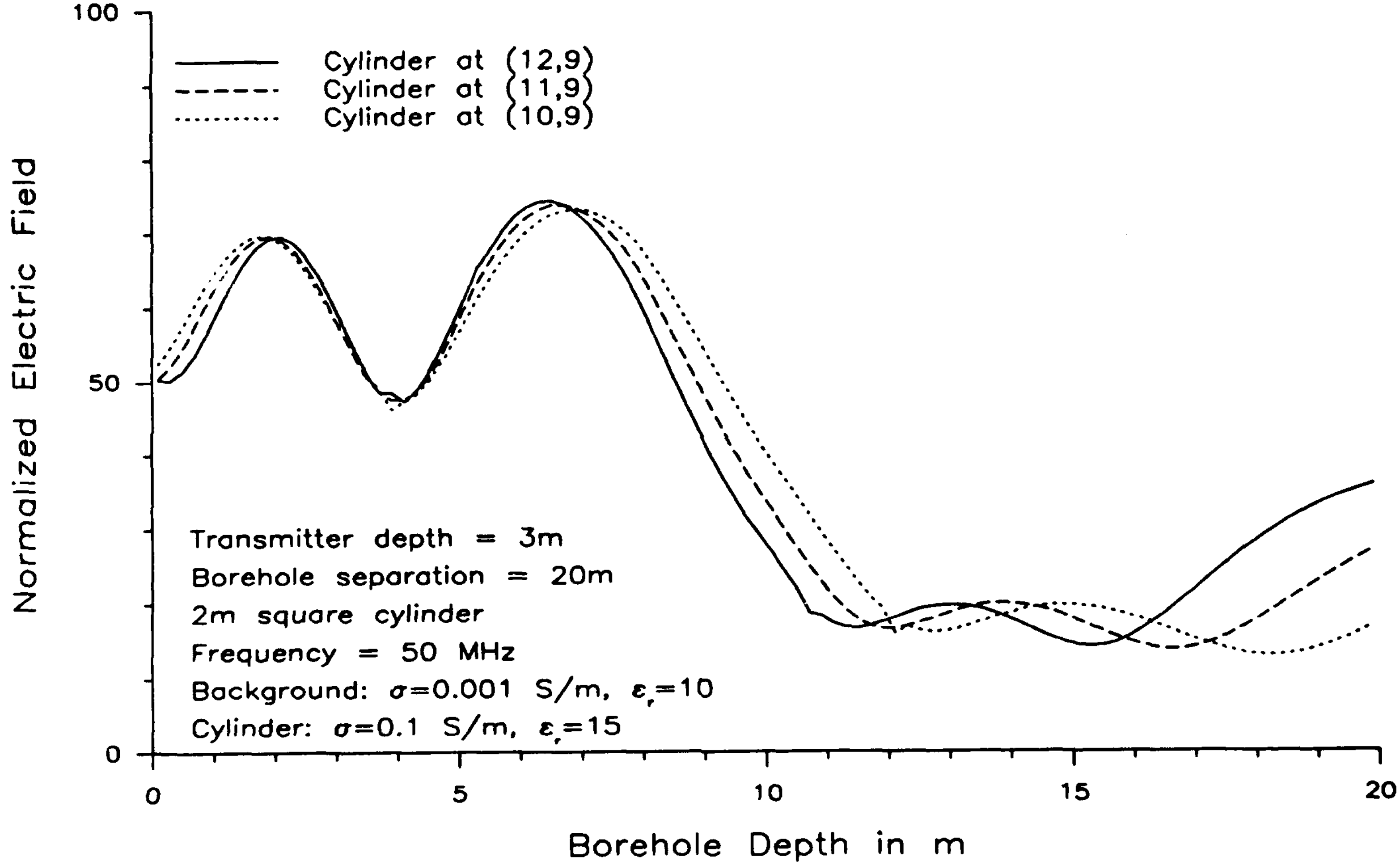


Fig. 3.4. Magnitude response of a conducting cylinder in a homogeneous earth for three different cylinder locations.

It is worth mentioning that this method does have a shortcoming, which is the same as that of the ray optics method discussed in the Chapter II. That is, the diffraction theory being used (refer to Chapter II) is based on the impedance boundary approximation. This approximation will be valid for higher conducting anomalies in a homogeneous earth, but will not be valid for lower conducting anomalies. However, the method will be applicable in those cases for which the approximation is a good one, and will be even more valuable if diffraction theory is extended to cover the lower conducting case.

3.2.4 Conclusions

In this section some methods for inverting geophysical data have been presented. The majority of these techniques were originally developed for medical imaging, and therefore might not be well suited for geophysical applications. In particular, the convolution-backprojection algorithm does not perform well when only an incomplete set of data is available, which is typically the case in geophysical tomography. The two main methods that do appear to work well in the geophysical setting are the algebraic method using the straight ray approximation, and diffraction tomography using either the Born or Rytov approximations.

The major limitation of the algebraic method was noted to be that the method ignored diffraction, reflection, and refraction effects, while diffraction tomography is only applicable for media containing slight inhomogeneities (that is, inhomogeneities whose electrical characteristics do not differ widely from the background).

To put these limitations into perspective, Figs. 3.5 and 3.6 show calculated electric field responses of a circular cylinder imbedded in a homogeneous medium. Each figure contains plots of the responses using the exact solution, the straight ray approximation, and the Born

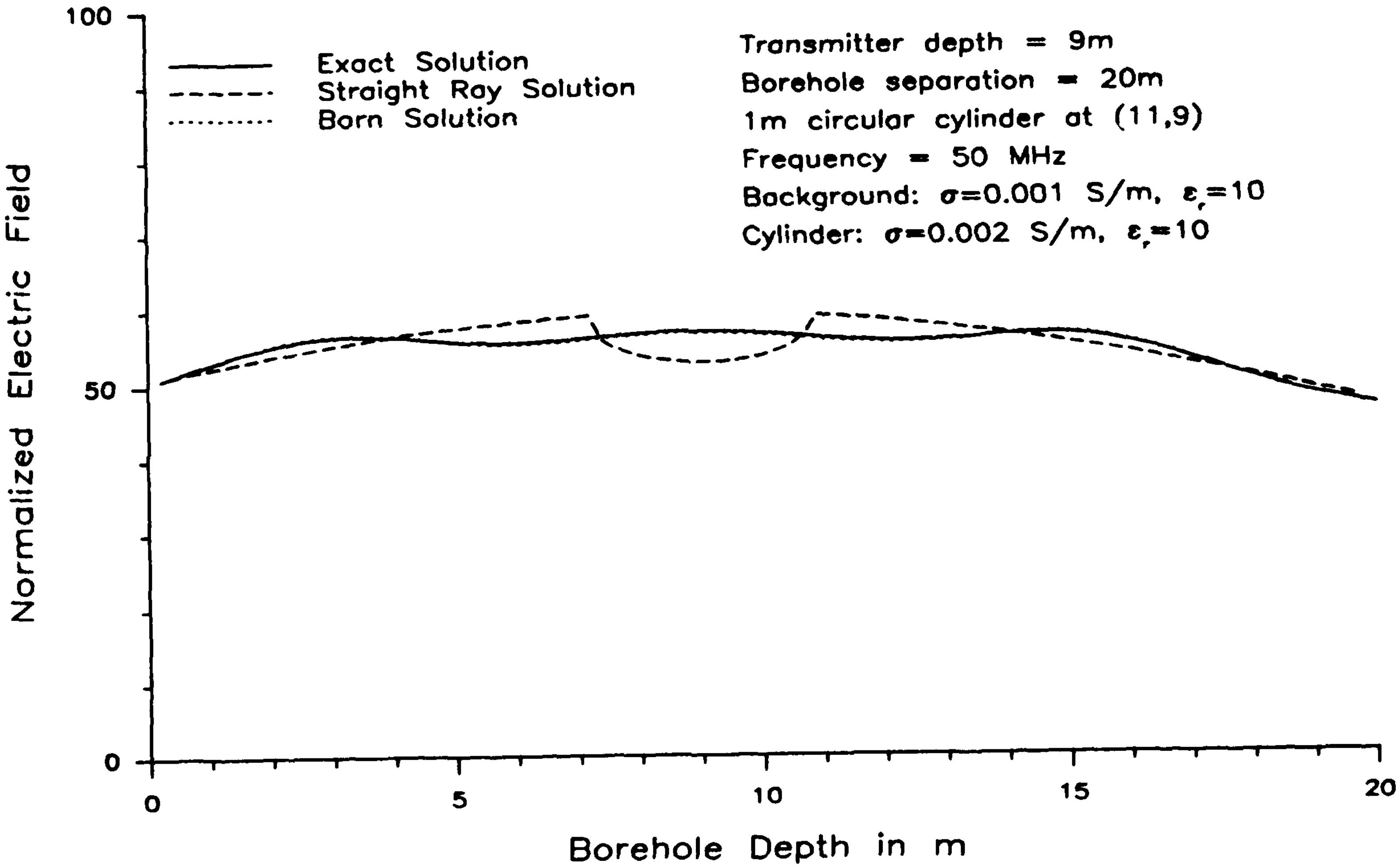


Fig. 3.5. Magnitude response of a slightly inhomogeneous cylinder in a homogeneous earth as a result of using straight ray approximation and Born approximation.

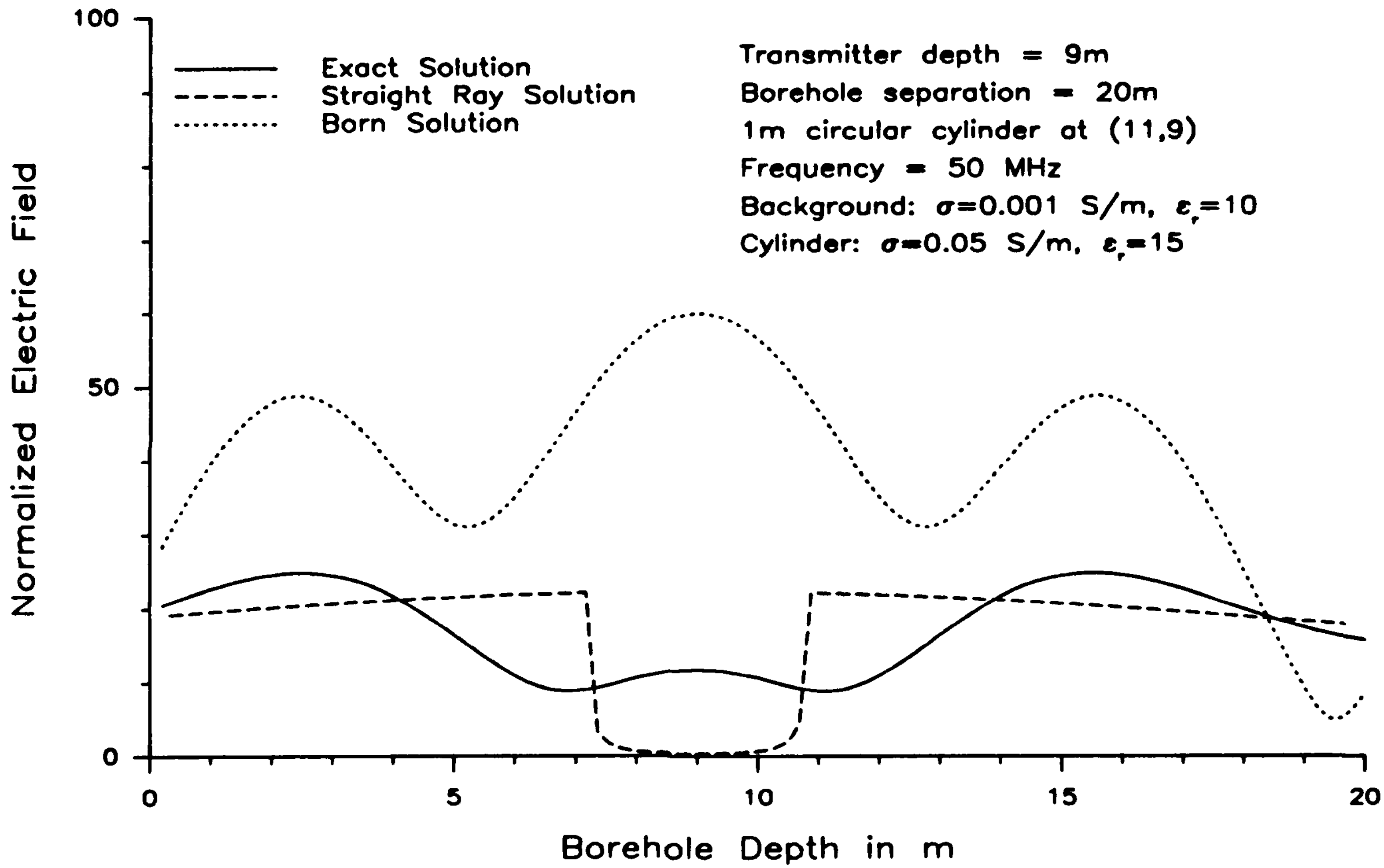


Fig. 3.6. Magnitude response of a strongly inhomogeneous cylinder in a homogeneous earth as a result of using straight ray approximation and Born approximation.

approximation. In both figures, the background medium is the same, while the cylinder characteristics are different. Fig. 3.5 is for the case of a slight inhomogeneity in that the conductivity is 0.002 S/m (the background conductivity is 0.001 S/m) and the relative permittivity is the same as the background. As can be seen from the data in this figure figure, the Born approximation matches the exact solution quite closely, while the straight ray solution is in error. For this case, one would expect that reconstructions using diffraction tomography would do very well. Fig. 3.6, on the other hand, is for a strongly scattering cylinder having a conductivity of 0.05 S/m and a relative permittivity of 15 (the background's is 10). As can be seen from this figure, the straight ray approximation underestimates the size of the null behind the cylinder, and overestimates the attenuation of the cylinder because it ignores diffraction of the fields. However, the Born approximation is badly in error for this case, since the scattered field is so large. For most cases, geophysicists are more concerned with detecting and locating large contrast anomalies (i.e., large scatterers), and ignoring slight inhomogeneities, therefore, it is felt that the algebraic method is the best inversion method for this application. For this reason, in the remainder of this dissertation the main emphasis will be on the algebraic method for geophysical inversions.

3.3 The Ill-Posed Nature of Geotomography

3.3.1 Introduction

There has been much interest in ill-posed problems, see for example, [45],[53],[54],[55],[6],[56]. The description of ill-posed problems was first given by Hadamard [6]. The definition of an ill-posed inverse problem is one that is not well posed. For the operator equation:

$$Ax = b \tag{3-39}$$

the problem of solving for the unknown vector x given data b and the mapping A (A is a matrix in the algebraic reconstruction method) is well posed if

a) An inverse (denoted by A^{-1}) exists for A, and

b) The inverse is continuous.

For the limited data tomography problem, an inverse does not exist. This is easily seen by considering the Radon operator given by (3-10) which measures the attenuation over the line L. This operator will have a non-zero kernel consisting of (at least) all two-dimensional functions (attenuations) which are defined in the region but are only non-zero off L. Recall that the kernel of an operator is the set of all vectors (functions) which are mapped by the operator to zero. The *Radon transform operator* (where all lines L are considered) has an inverse which was found previously using the projection theorem. However, the inverse is not continuous when the mapping (A) is from the set of square integrable functions on the region ($L_2([a,b]\times[a,b])$) into square integrable functions on the unit cylinder [57]. The authors in [57] do show that the inverse is continuous when the mapping is taken to be between Sobolev spaces.

So the tomography problem in its original form is ill-posed. Our main interest here is in the algebraic reconstruction method for which the mapping, A, is an $m\times n$ matrix, where m is the total number of measurements and n is the number of pixels in the reconstructed image. Since it is assumed that $m>n$ (an overdetermined system of equations), A does not have an inverse. We can, however, consider a generalized inverse which when post-multiplied by the measurement vector results in a least squares solution of (3-39) (See Appendix B for a discussion of least squares solutions). This generalized inverse has the form

$$A^+ = (A^TA)^{-1}A^T, \tag{3-40}$$

such that

$$J(x) = \|AA^+x - b\| \tag{3-41}$$

is a minimum. In this dissertation, $\|\circ\|$ refers to the Euclidean norm unless otherwise stated (see Appendix B). The matrix A^+ is obviously continuous if the inverse on the right side exists (i.e., the rank of A = n). However, small errors in the measurement vector may imply large errors in the least squares solution as can be seen from theorem 5.2.4 in Stewart [58]

$$\frac{\|A^+b - A^+\hat{b}\|}{\|A^+b\|_2} \le \kappa(A) \frac{\|b_1 - \hat{b}_1\|}{\|b_1\|} \tag{3-42}$$

where

$$\kappa(A) = \|A\| \, \|A^+\|. \tag{3-43}$$

The vector $\hat{b}$ is the measurement vector with added error and b_1 and $\hat{b}_1$ are the projections of b and $\hat{b}$ onto the range of A. What this theorem says is that the error in the solution is bounded by some measure of the error in the measurement vector times the condition number of the matrix. If the matrix is nearly singular (i.e., close to some singular matrix), then the condition number is large and the equation can be considered to be in some sense ill-posed. However, numerical analysts usually refer to the system as being ill-conditioned.

Unfortunately, the situation is even worse for cross-hole tomography in that the matrix A does not even have full column rank. In fact an upper bound on the rank is given by [59]

$$\operatorname{rank}(A) \leq n - \tfrac{1}{2}(N_{VZ} - 1), \tag{3-44}$$

where, N_{VZ}, the number of vertical zones, is the number of pixels in each row of the discretized image (see Fig. 3.2). A generalized solution still exists which can be obtained using the singular value decomposition (see appendix B). In any case the problem is ill-posed, so we discuss the method of regularization for better posing the problem.

3.3.2 Regularization

A. Introduction

Regularization consists of methods for solving a problem similar to the original inverse problem given by (3-39), but with solutions which are not as sensitive to errors in the data. In fact, the method of least squares mentioned above is one such technique in that it gives an alternate description of the solution. In addition, for an inconsistent system of equations there is no x such that $Ax=b$, and in this case we are restricted to finding an x which minimizes $\|Ax-b\|$. Methods also exist for restricting the space in which solutions are to be sought. This might come about by enforcing constraints that the solution has to satisfy (e.g., a positivity constraint by which all pixels in a solution image are required to have positive value).

In this section the method of regularization will be restricted to methods of finding a set of approximate solvers for (3-41) [50], [53], [54], [55], [56], [60]. What is desired, is a solution of the form

$$x = A^{+} b. \tag{3-45}$$

Since A^+ does not exist, we attempt to find A_γ such that

$$x_\gamma = A_\gamma b, \tag{3-46}$$

and

$$\lim_{\gamma \to 0} x_\gamma = x \tag{3-47}$$

where γ is known as the regularization parameter. Unless otherwise indicated, by a solution we mean a least squares solution in the following. In general the limit in (3-47) does not really exist because A_γ will not exist as $\gamma \to 0$. The goal is to choose γ sufficiently small to get a good approximation to the solution, yet sufficiently large such that A_γ remains bounded and does not magnify errors in b.

Three methods of regularization will be discussed:

1) Tikhonov regularization,
2) the truncated singular value decomposition, and
3) truncated iterative procedures.

It will be seen, however, that these methods are related. Although these different techniques will be described for an $m \times n$ matrix, they are applicable to a general linear operator between Hilbert spaces. This setting might be important if, for example, discretization were to be performed after regularization (see [60] for further details).

B. Tikhonov Regularization

As previously stated, our interest is in solving (3-41). That is, find an approximate solution, x_{LS}, such that $\|Ax_{LS}-b\|^2$ is a minimum (see appendix B for a more detailed discussion). As was previously noted, this problem can be very ill-conditioned in that small errors in

b can lead to large errors in x_{LS}. Tikhonov [54] suggested trying to minimize the alternative functional

$$J(x) = \|Ax-b\| + \gamma\|Mx\|, \qquad (3\text{-}48)$$

where M is an n×n matrix which we will take to be the identity matrix in the following. J(x) can be minimized with respect to x by taking its derivative with respect to x as

$$\frac{\partial J(x)}{\partial x} = \frac{\partial}{\partial x} \langle Ax-b, Ax-b\rangle + \gamma\langle x, x\rangle$$

$$= -2b^T A + 2x^T A^T A + 2\gamma x^T, \qquad (3\text{-}49)$$

and then setting the derivative equal to zero

$$-b^T A + x^T A^T A + \gamma x_\gamma^T = 0$$

$$x_\gamma^T (A^T A + \gamma I) = b^T A$$

$$x_\gamma^T = b^T A (A^T A + \gamma I)^{-1}$$

$$x_\gamma = (A^T A + \gamma I)^{-1} A^T b, \qquad (3\text{-}50)$$

where I is the n×n identity matrix. Equation (3-50) is also referred to as ridge regression or the Lagrange multiplier method. The problem then reduces to optimally choosing the regularization parameter γ. Some methods for choosing γ will be discussed later.

C. *Truncated Singular Value Decomposition (SVD)*

The SVD is a well known expansion for an operator (e.g., a matrix) [61], (see also appendix B). The decomposition for the matrix A is given by

$$A = U\,S\,V^T, \tag{3-51}$$

where U (m×m) and V (n×n) are orthogonal matrices, and S (m×n) has the form

$$S = \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} \tag{3-52}$$

where the 0 represents a matrix having all zero elements and of dimension (m-n) by n, and Σ is a diagonal matrix (of dimension n) whose elements (called singular values)

are the square roots of the eigenvalues of A^TA. Using this decomposition, the original equation can be transformed into an uncoupled system of equations as [45]

$$\begin{aligned} Ax &= b + e \\ U\,S\,V^Tx &= b + e \\ S\,V^Tx &= U^Tb + U^Te \\ Sx' &= b' + e', \end{aligned} \tag{3-53}$$

where the error in the system has been represented by e and x' and e' are the representations of x and b in the new coordinates. In deriving the above, we have used the fact that, $U^TU = I$. Since S has the form given in (3-52), it is easy to solve for x' as

$$\xi_i' = \frac{1}{\sigma_i}(\beta_i' + \varepsilon_i'), \tag{3-54}$$

where the σ_i are arranged in descending order. It is assumed that e'

is a random vector with all of its elements having similar magnitudes. We also assume that the elements of x' have similar sizes. This may be an unfair assumption, and actually it is desireable to have the largest components of x' correspond to the largest singular values. With these two assumptions in mind, it can be seen that the β_i' decreases with i, and at some point the ε_i' will dominate in (3-54). It is at this point that the SVD is truncated as

$$\xi_i' = \begin{cases} \frac{1}{\sigma_i}(\beta_i' + \varepsilon_i'), & \text{for } i \leq N \\ 0 & , \text{for } i > N \end{cases} \tag{3-55}$$

where the integer 'N' is chosen as suggested above. Transforming back to the original coordinates, the following formula for the solution can be obtained

$$x = \sum_{i=1}^{N} \frac{1}{\sigma_i} u_i^T(b+e)\, v_i, \tag{3-56}$$

where u_i is the i^{th} row of U, and v_i is the i^{th} row of V. Note that 'N' must be less than the rank of A, or a divide exception may occur. Tikhonov regularization may be used with the SVD to obtain a formula similar to (3-56) (see Appendix B for details)

$$x = \sum_{i=1}^{N} \frac{\sigma_i}{\sigma_i^2 + \gamma} u_i^T(b+e)\, v_i, \tag{3-57}$$

Note that γ has effectively damped the effects of the small singular values, consequently the name 'damped least squares' has also been applied to Tikhonov regularization.

D. Truncated Iterative Procedures

It is a well-known phenomena that some iterative procedures which find a solution to (3-41) converge to a good solution and then diverge from this solution. This behavior has been observed in the algebraic reconstruction technique [59], the method of conjugate gradients, and the method of successive approximations [55]. It appears that in a fashion similar to the SVD, the noise tends to dominate as the solution progresses.

In a general iterative scheme, the solution at the (k+1)st step is given by

$$x_{k+1} = f(A,b,x_k),\ k = 1,N \tag{3-58}$$

where f(∘) represents the particular iterative algorithm and N is the number of iterations to be performed. The regularization simply consists of choosing N, the number of iteration steps sufficiently small to minimize the effects of error in the measurement data, yet sufficientl large to get a good approximation to the solution. Methods for choosing N will depend on the iterative algorithm, and will be discussed later.

3.3.3 Constrained Solutions

The geotomography problem has been noted to be ill-posed, which in this setting means that the solution does not depend continuously on the data. For example, one could find a set of solutions to the problem which do depend continuously on the data, then it would be best to search in this set for the unknown image. This would be one method of constraining the solution. In fact, most regularization schemes attempt to accomplish this goal.

In addition, we can also constrain the solution based on our

underlying knowledge of the problem. For example, we might want to search for (solution) images having attenuation values within a certain range, or parts of an image to have attenuation values equal to some previously known value. In this way, the resulting image will one that satisfies the physical constraints imposed upon it. The imposition of these physical constraints may also produce a regularizing effect on the solution [60]. However, our primary interest is in imposing the constraints in order to obtain a physically realizable solution.

CHAPTER IV

NUMERICAL ALGORITHMS FOR IMAGE RECONSTRUCTIONS

4.1 Introduction

In this chapter we will develop algorithms for solving the system of equations arising from the discretized (image) tomography model. As discussed in Chapter III, the first step in regularizing this problem is to search for a least squares solution to the problem. However, for the case at hand, even a least squares solution will be subject to large variations as result of relatively small changes (often due to noise) in the measurement data. Therefore, it will be important to insure that the algorithms developed are relatively insensitive to noise in the data. Finally, the algorithms must be able to incorporate *a priori* information in the form of inequality constraints on the solution. This will in general provide a further regularizing effect on the solution.

Although computational requirements of the algorithms are a concern, the computations are usually performed 'off-line', so speed is not a major consideration. However, the solution image vectors may be composed of a large number of pixels, so that algorithms which can operate 'out-of-core' are desireable.

4.2 Least Squares Solutions

4.2.1 Introduction

As stated in the Chapter III we consider the problem of finding a solution vector x which minimizes $\|Ax-b\|$ as a means of solving the discretized tomography problem, given by equation (3-26). For a general description of least squares solutions, see Appendix B. The name "least squares" comes from the fact that the norm being used is

the euclidean norm. This norm is given by the square root of the sum of the squares of the elements of a vector. The minimization could be performed using some other norm such as the 1-norm or ∞-norm, but the 2-norm or euclidean norm has some important theoretical and computational advantages. Some of these advantages are listed below

a) The 2-norm has an associated inner product (i.e., $\|x\|^2 = \langle x,x\rangle$), which adds a geometric structure to the problem. In addition, the problem can be solved by setting to zero the derivative of $\langle (Ax-b),(Ax-b)\rangle$ resulting in the consistent normal equations (see appendix B) which are sometimes easier to solve. This inner product (which is not available with 1-norm or ∞-norm minimizations) affords additional computational advantages. See [62] for a discussion of these advantages and an overview of least squares methods.

b) The least squares solution arises naturally when a maximum likelihood estimator is used to solve (3-26) [63]. The maximum likelihood estimate is the one which maximizes the conditional probability density function $p(b|x)$. The vector b is the data vector, and x is the unknown model vector. If the assumption is made that the error vector, e, in (3-26) is zero mean Gaussian distributed, then the maximum likelihood estimate of x is given by [63]

$$\hat{x} = (A^T R_e^{-1} A)^{-1} A^T R_e^{-1} b, \tag{4-1}$$

where R_e is the covariance matrix for the random vector e. If it is further assumed that the elements of e are uncorrelated, and have identical variances σ^2, then (4-1) reduces to

$$\hat{x} = (A^T A)^{-1} A^T b. \tag{4-2}$$

Note that (4-2) gives the least squares solution via the generalized inverse, (3-40).

Therefore, from the discussion in the Chapter III on regularization, we are led to solving the approximate problem of minimizing the norm of the residual. In addition, from the reasons cited above, the norm we will use is the euclidean norm which results in a least squares solution.

The least squares problem for limited data reconstructions has some particular requirements which may not exist in other settings. These requirements need to be kept in mind when developing algorithms for solving the reconstruction problem. Some of these requirements are listed in the following.

a) The reconstruction problem often leads to large matrices and vectors. For example, a cross-hole arrangement for scanning a region 20 meters on a side will require an image vector having 400 elements for a *1 meter pixel size*. The associated distance matrix will have 400 columns and over 400 rows (for an overdetermined system of equations). For configurations resulting in such a large distance matrix, it may not be feasible to store the matrix in core memory. This requirement would preclude the use of inversion algorithms which operate directly on the distance matrix.

b) In addition to being large, the distance matrix often has many elements equal to zero. This 'sparseness' is due to the fact that each ray will intersect only a small percentage of the pixels in the region. It is important to develop algorithms which can take advantage of this sparseness to reduce computer memory and time requirements.

c) It will be necessary to incorporate the theory of regularization, discussed in the Chapter III, into any least squares algorithm developed.

d) It is often necessary to constrain the solution vector (i.e.,

image) to have non-negative values, or values in some specified range.

All of these requirements will be considered when designing the image reconstruction algorithms in this chapter.

4.2.2 Alternate Descriptions of the Least Squares Problem

A. Introduction

Because of the geometry involved in the least squares solution, there exist other systems of equations which when solved also yield a least squares solution. These alternate descriptions are important for the computational advantages that they possess. We list some of the other descriptions and point out their computational advantages.

B. Normal Equations

The normal equations are derived in Appendix B. They are repeated here in matrix form for convenience

$$A^TAx = A^Tb. \tag{4-3}$$

This equation is the result of requiring that the residual vector of a least squares solution is orthogonal to the columns of A. This equation has a number of computational advantages over solving the original equation, $Ax=b$. The advantages are listed below.

1) If round-off errors are neglected, (4-3) represents a consistent set of equations. This is especially important when using the projection method [64] (also known as the ART algorithm) for finding a solution since this algorithm only converges when the set of hyperplanes given by the individual equations intersect at a single point. This fact will be

discussed later when the projection method is fully described.

2) Equation (4-3) has n equations with n unknowns. For m>n (an overdetermined set of equations), (4-3) will require less computational labor than the original equation.

3) The coefficient matrix (A^TA) in (4-3) is symmetric, positive semidefinite. This is often a requirement in using gradient algorithms for finding a least squares solution.

The normal equations do have two disadvantages that are not so deleterious if they are given proper attention. These disadvantages are:

1) It is well-known that forming the normal equations results in a squaring of the condition number of the system of equations [58]. This follows easily from

$$\kappa(A^TA) = \kappa(A)\kappa(A) \tag{4-4}$$

where $\kappa(\circ)$ is the condition number of the given matrix as defined in Appendix B. As Stewart [58] points out,this problem may be alleviated by performing the computations in double precision.

2) Forming the normal equations will, in general, result in a loss of sparsity in the coefficient matrix. This will be a problem when the size of the problem is sufficiently large that it is more efficient to use the sparseness to reduce storage and time requirements. This problem may be avoided by using an iterative scheme where A^TA does not have to be explicitly formed.

The normal equations should be considered as a viable method of solving the least squares problem. However, the disadvantages listed above should be kept in mind when using this technique.

C. *Iterative Refinement*

The following augmented matrix equation was studied by Bjorck [65] as a method of improving a least squares solution

$$\begin{bmatrix} 0 & A^T \\ A & I \end{bmatrix} \begin{bmatrix} x \\ r \end{bmatrix} = \begin{bmatrix} 0 \\ b \end{bmatrix}. \tag{4-5}$$

However, it may be useful to solve (4-5) directly for the least squares solution. This method is similar to the normal equations in that the coefficient matrix is symmetric, and that use is made of the fact that the residual vector is orthogonal to the columns of A. Note that the new coefficient matrix in (4-5) need not be explicitly formed if an iterative procedure is used to solve (4-5). However, using (4-5) has the disadvantage of requiring the solution of a larger set of equations. Using (4-5) may also have the same conditioning problem as the normal equations. Bjorck [65] has shown that the condition number of the new coefficient matrix lies between the condition of A and the square of the condition of A. If for a particular geometry the condition number can be determined to be lower than A^TA, then (4-5) may be used in lieu of the normal equations.

4.2.3 Weighted Least Squares

A. Introduction

When comparing the maximum likelihood estimate and the least squares solution at the beginning of this chapter, a noise covariance term was used. This term was eliminated from (4-1) by assuming that the elements of the noise vector are uncorrelated and have equal variances. We now drop the assumption that the elements have equal variances. Rather, it is assumed that there exists *a priori* knowledge that some of the equations in the model are more reliable than others.

Two practical examples of this *a priori* knowledge will be presented in this section. In this case, the inverse of the noise covariance matrix is replaced by a weighting matrix W. This matrix has the form

$$W = \begin{bmatrix} \omega_1 & & 0 \\ & \ddots & \\ 0 & & \omega_m \end{bmatrix}, \qquad (4\text{-}6)$$

where the scalar ω_i will weight the equation corresponding to the i^{th} ray path measurement. The least squares minimization problem now involves finding the solution x_{WLS} such that the functional

$$J_{WLS}(x) = \|W(Ax_{WLS} - b)\|_2 \qquad (4\text{-}7)$$

is a minimum. Note that if there is a unique solution to (4-7) [i.e., $b \varepsilon R(A)$], then $x_{LS} = x_{WLS}$. Of course, it would be overly optimistic to hope for this to happen in practice. The weighted normal equation obtained from (4-1) is found to be

$$A^T WAx = A^T Wb \qquad (4\text{-}8)$$

B. Path Length Weighting

We now address the problem of determining the diagonal elements of the matrix W. One method often suggested is to make ω_i inversely proportional to the standard deviation of the i^{th} measurement. This process would obviously give greater weight to the more reliable measurements. The major drawback to this method is that the statistics of the measurement data are not usually available. We suggest a more elementary means of weighting by letting

$$\omega_i = d_i^{-\alpha} \tag{4-9}$$

where d_i is the distance the i^{th} ray travels from the transmitting antenna to the receiving antenna, and α is some positive constant which will be dependent on the geometry of the problem and the amount of noise in the measurement data. Note that a type of path length weighting was also used in [11] in conjuction with the algebraic reconstruction technique (ART), but using the weighting method (4-6) is more flexible since we are not restricted to single inversion algorithm. On the average, the received signals resulting from shorter paths have a greater signal to noise ratio than those from longer paths. Thus, this type of weighting is attractive since it is expected that measurements resulting from shorter paths will be more reliable.

In practice, it is more useful to normalize d_i by dividing it by the shortest possible distance, d, to get

$$\omega_i = \left[\frac{d_i}{d}\right]^{-\alpha} \tag{4-10}$$

This WLS method is easy to implement on a computer since the A matrix contains the path length information. In fact, d_i is found by summing all of the elements in the i^{th} row of A. As noted above, the constant α should be chosen with consideration to the amount of noise in the measurement data. To understand this relationship, the effects of adding noise will be investigated in more detail later.

C. *Estimated Received Power Weighting*

It has been shown [66] that the geotomography problem can be better posed by calculating estimated received powers. The procedure

is to use a preliminary estimate of the attenuation over the region (i.e., the image) and substitute this into the forward equation given by (3-6). The system of equations is then modified by eliminating those equations whose measured received power differs significantly from the estimated power. The criterion used in [66] was to select a constant k such that if

$$\begin{array}{ll} k\, P_{rec_i} \geq P^*_{rec_i} \geq \frac{1}{k} P_{rec_i} & \text{path is retained} \\ \text{otherwise} & \text{path is eliminated} \end{array} \tag{4-11}$$

where P_{rec_i} is the measured power for the i^{th} path and $P^*_{rec_i}$ is the estimated received power. This procedure has the advantage of removing equations which may not accurately model the geotomography problem, but has the disadvantage that in removing equations (i.e. rows of A) the rank of A may be decreased. We propose here to retain all paths, but inversely weight those paths which are suspect according to (4-11). Applying these results, a new W matrix is obtained with

$$W = \begin{bmatrix} \omega_1\xi_1 & & & & \\ & \ddots & & & \\ & & \omega_i\xi_i & & \\ & & & \ddots & \\ & & & & \omega_p\xi_p \end{bmatrix} \tag{4-12}$$

where ω_i is chosen as in (4-10) and ξ_i is given by

$$\xi_i = 1 - \frac{\left| P_{rec_i} - P^*_{rec_i} \right|}{P_{rec_i}} \tag{4-13}$$

if the path met the elimination criterion in (4-11). Otherwise, ξ_i is set to unity so that the weighting is unaffected by this procedure.

It is seen that solving the underground reconstruction problem

using weighted least squares is a four step process:

1) The weighting matrix is formed as in (4-6) using path length weighting.
2) The weighted normal equation (4-8) is solved using some least squares algorithm to find a cross sectional image, x_{WLS}.
3) This cross sectional image is then used to calculate estimated received powers and a new weighting matrix (4-12).
4) Equation (4-8) is now solved using the new W matrix resulting in an improved cross sectional image.

D. Illustration of the Effects of Additive Noise

If the assumption is made that the noise is additive, white and Gaussian (AWGN), then a signal-to-noise ratio can be defined after Ney, ***et al.*** [67] as

$$SNR = 20\log_{10}\left\{\frac{\|E^i\|}{\|n\|}\right\}, \tag{4-14}$$

where E^i is the vector whose elements are the magnitudes of the electric fields incident at the receiving antennas, and n is a normally distributed pseudo-random noise vector.

Fig. 4.1 shows the predicted electric field magnitude, the electric field with additive noise, and the random noise vector, all plotted versus borehole depth (in the receiving borehole) for a cross-hole configuration with a homogeneous earth. The signal-to-noise ratio was chosen to be equal to 30 dB. As can be seen from the plot, the additive noise is less significant at those receiver locations directly opposite the transmitting antenna than for those locations near the top or bottom of the borehole.

This example gives justification for using the WLS method in that those paths with the shorter ray distance (from a given transmitter to a given receiver) will on the average have the higher signal-to-noise

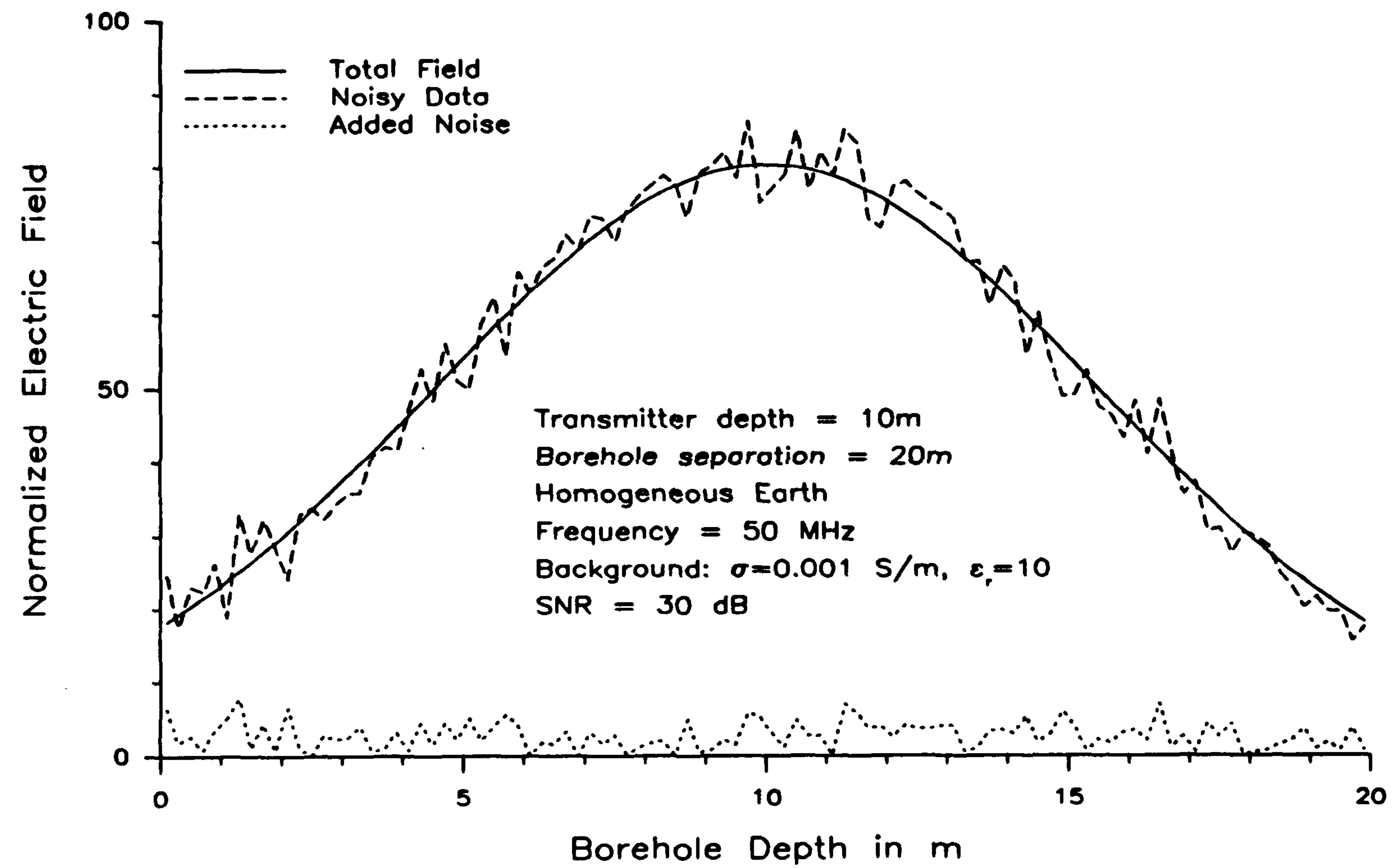


Fig. 4.1. Magnitude response of a homogeneous earth showing the effects of adding random noise to the measurement data.

ratio. These 'reliable' paths will then be given stronger weighting, and therefore will have greater influence on the reconstruction.

Also, note that as the amount of noise increases, the measurements resulting from the longer ray paths will become less reliable when compared to the shorter ray path measurements. Therefore, for applications where the magnitude of the additive noise is expected to be large, one should use a greater value of α in (4-10) than in cases where the noise is not expected to be significant. However, α can not be made too large, or resolution in the horizontal direction will be compromised since the longer ray paths give horizontal information. Also, some consideration should be given to the effect α has on the conditioning of the system as shown in the next section.

E. Numerical Considerations for Path Weighting

The WLS method is easily applied to geophysical inversion problems. The A matrix contains the path length information, so the weighting matrix, W, is easy to generate. Once the W matrix has been found, the matrix multiplications given in (4-8) can be carried out to get a new matrix equation of the form

$$Cx = d \tag{4-15}$$

where C is a symmetric $n \times n$ matrix and d is an $n \times 1$ vector. Certainly these matrix multiplications are additional computations which would not normally be required, but since in most cases $m>n$, (4-15) represents a reduced set of equations over that of solving $Ax=b$, and will therefore require fewer calculations.

We would also like to consider the conditioning of the system of equations representing the reconstruction problem. The equations to consider are (4-3) for the least squares problem (using the normal equations) and (4-8) for the weighted least squares problem. For

either problem, the reduced equation is of the form given in (4-15), where $C = A^T A$ for the least squares problem and $C = A^T WA$ for the weighted least squares problem. The condition number for (4-15) represents the susceptibility of the solution to errors (noise) in the data. The condition number is defined in Appendix B.

Table 4.1 shows the effect weighting has on the conditioning of the system of equations. For this table the matrix A was formed by considering an underground reconstruction problem consisting of 20 transmitting and 20 receiving antennas, and the cross sectional image divided up into 100 square cells. The table demonstrates the effect the exponential factor, α in (4-10), has on the condition number. As can be seen from the table for values of α between 0.5 and 3.0, the rank of $C = A^T WA$ is actually greater than the rank of $C = A^T A$. This increase in rank results in a system of equations with greater stability. Also from the table it can be seen that the optimum value of α from a conditioning standpoint is between 1 and 2.

Table 4.1
Choosing the path weighting exponent

α	rank LS	κ_{LS}	rank WLS	κ_{WLS}
0.5	93	22×10^3	93	12×10^3
1.0	"	"	94	12×10^3
2.0	"	"	95	17×10^3
3.0	"	"	93	8×10^3

4.2.4 Conclusions

In this section a discussion of the applicability of least squares methods to the image reconstruction problem has been presented. Along with this discussion, some alternate methods for

solving the least squares problem have been reviewed. A generalization of the least squares solution (i.e., weighted least squares technique) has also been reviewed. Two new methods for determining weighting coefficients for the WLS technique have been introduced. These weighting coefficients are based on the following.

a) Making the observation that the geometry of the cross-hole configuration makes some measurements inherently more reliable than others, and

b) Using an initial estimation of the cross section to determine the reliability of each measurement.

Now that the least squares problem has been described, and a means of improving the least squares solution by weighting the measurements has been suggested, in the remainder of this chapter numerical algorithms will be introduced which solve the LS and/or WLS problems.

4.3 A Direct Algorithm Using the Singular Value Decomposition

4.3.1 Introduction

Direct algorithms for finding a solution to $Ax=b$ operate directly on the matrix A in solving for x. The most well-known direct algorithm is Gaussian elimination which attempts to reduce A to upper trapezoidal form. Once A has this form, a solution is easily obtained. Unfortunately, Gaussian elimination will not work when A is rank deficient or ill-conditioned. Since we are faced with a rank deficient matrix in this setting, some other more stable algorithm must be used. The singular value decomposition is one such algorithm which will be developed in this section. Not only is this decomposition useful in its own right, but it also provides analytical tools for studying the properties of other algorithms which will be presented later in this chapter.

4.3.2 Singular Value Decomposition Algorithm

The singular value decomposition (SVD) was introduced in Chapter III as a means of regularization for an ill-posed problem. It is also discussed in Appendix B as a way of computing the generalized inverse of a matrix. The solution using the SVD was given in (3-57), it is repeated here for convenience

$$\hat{x}_\gamma = \sum_{i=1}^{N} \frac{\sigma_i}{\sigma_i^2 + \gamma} u_i^T b \, v_i , \qquad (4\text{-}16)$$

where γ is the regularization parameter, and the error vector, e, has been dropped for simplicity. There are three issues that must be addressed before using (4-16).

a) What value should be chosen for N?

b) What value should be chosen for γ?

c) How can constraints be applied to the solution?

These issues will be given consideration in the succeeding sections.

4.3.3 Truncating the Singular Value Decomposition

In the last chapter it was noted that the SVD was able to effect a change of coordinates such that in the new coordinate system, the equation Ax=b was reduced to diagonal form. Once reduced, the equation is easily solved. However, as noted in (3-54), errors in the elements of the data vector corresponding to small singular values can cause large errors in the solution. The goal is to choose N in order to ignore the terms which cause large errors. In Fig. 4.2 are plotted the singular values, elements of the data vector, and the ratio of these two quantities for a system of equations arising from a discretized cross-hole configuration. The singular values are plotted in decreasing order. As can be seen from the figure, the ratio β/σ stays

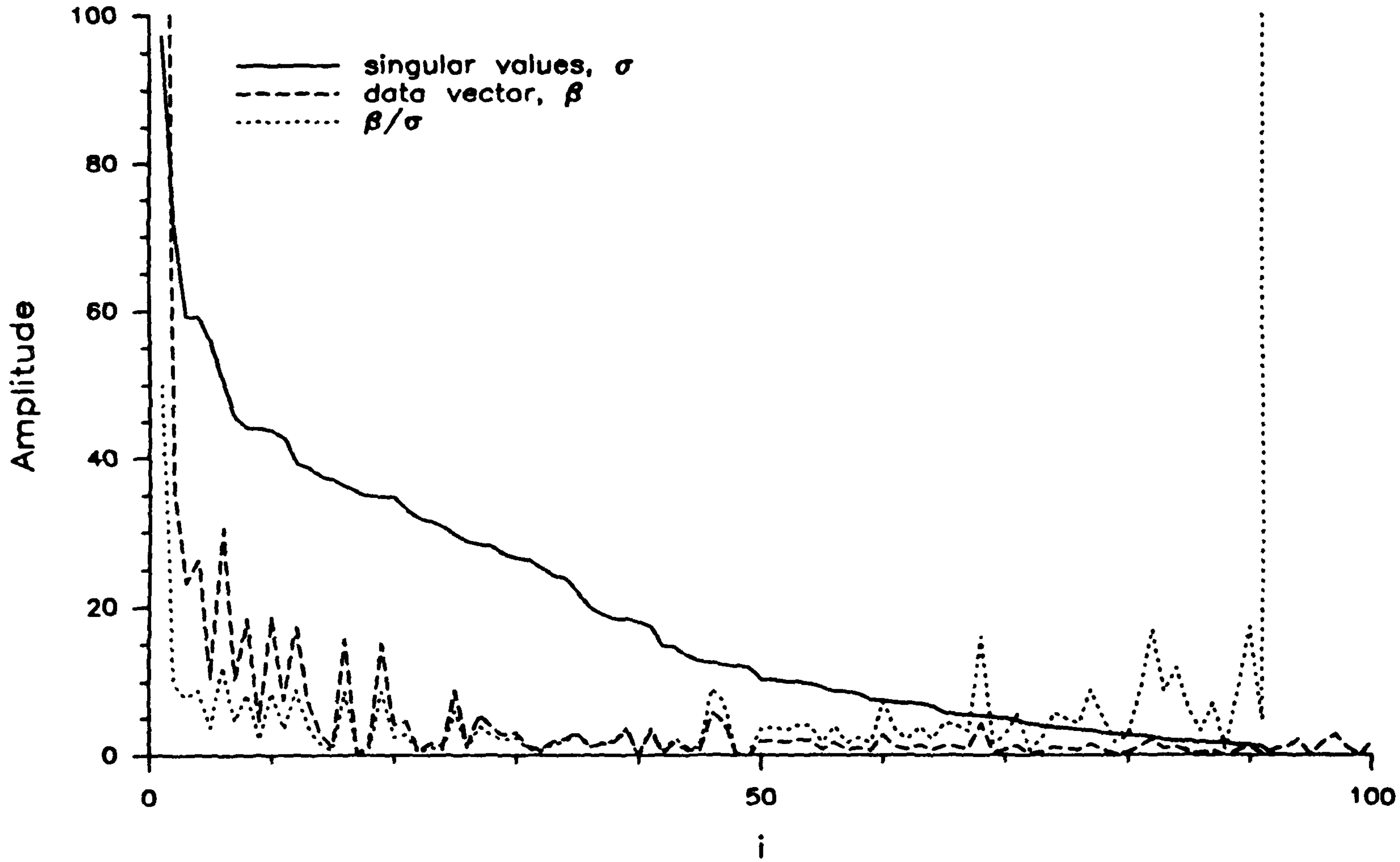


Fig. 4.2. Singular values, transformed measurement vector components, and their ratio for the set of equations resulting from a cross-hole measurement system.

relatively constant, and then takes a large jump at i=91. If we assume that the components of the solution have similar magnitudes in the new coordinate system, then it appears that the error in the data vector dominates the real data for i>90. In this case, it would make sense to choose N=90.

In general, when solving a system of equations the ratio β/σ can be averaged, and a large deviation from the average could be detected. At this point, N would be chosen to be one less than the value of i at which the large deviation was found.

4.3.4 Tikhonov Regularization for the SVD Algorithm

This regularization method was described in the Chapter III. The method is given by (4-16), with N = n which is the number of elements in the x vector. In this section some ways of choosing a value for γ in (4-16) will be discussed.

One method for choosing γ is taken from statistical analysis [68], and , it is based on observing the elements of the solution vector (i.e., x) as γ is increased from zero. A plot of these elements versus γ is referred to as a ridge trace. The selection process involves finding the smallest value of γ after which the ridge trace does not fluctuate. This value of γ is then used in (4-16). Unfortunately, this method has the disadvantage of requiring that (4-16) be solved for different values of γ. In addition, a subjective test is needed to choose the minimum value of γ.

Another method using statistical analysis is the generalized cross-validation method which attempts to minimize the expected value of the error between the actual solution and the regularized solution [69]. In this case, the expected value function has to be calculated versus γ so that a minimum can be found.

A more efficient method of choosing γ would be in line with the discussion above on truncating the SVD. That is, detect the point at

which the error dominates in the summation of (4-16), and then choose γ to damp out the effects of the small singular values. For example, if it has been found that the error is dominant for $i>90$ in (4-16) then choose γ on the order of σ_{90}^2. This procedure should introduce a small bias into the lower order terms while reducing the effects of noise in the higher order terms. In Fig. 4.3, β_i/σ_i is again plotted with singular values damped as

$$\sigma_{\gamma i} = \frac{(\sigma_i^2 + \gamma)}{\sigma_i}. \tag{4-17}$$

Also plotted is the undamped version of β_i/σ_i. As can be seen from the figure, the damping has reduced the error in the high order terms, while leaving the low order terms relatively unaffected. This method has an advantage over the truncation method in that the u_i coordinates for large i are not neglected. However, the magnitudes will be reduced along these axes. In addition, if the noise in the measurement vector increases, such that it starts to affect some of the lower order terms, then this method will be more immune to the additional noise.

4.3.5 Applying Constraints in the SVD Algorithm

The most obvious method of applying constraints would be to project the solution given by (4-16) onto the constraint surface. For example, if a nonnegativity constraint is to be applied, then set all of the negative elements in the solution to zero. Unfortunately, this procedure does not result in good reconstructions, since it has been observed by the author that the resulting images have a significant number of pixels which are zero. One would expect, in general, that the images should not have any zero-valued pixels. Rather, the reason

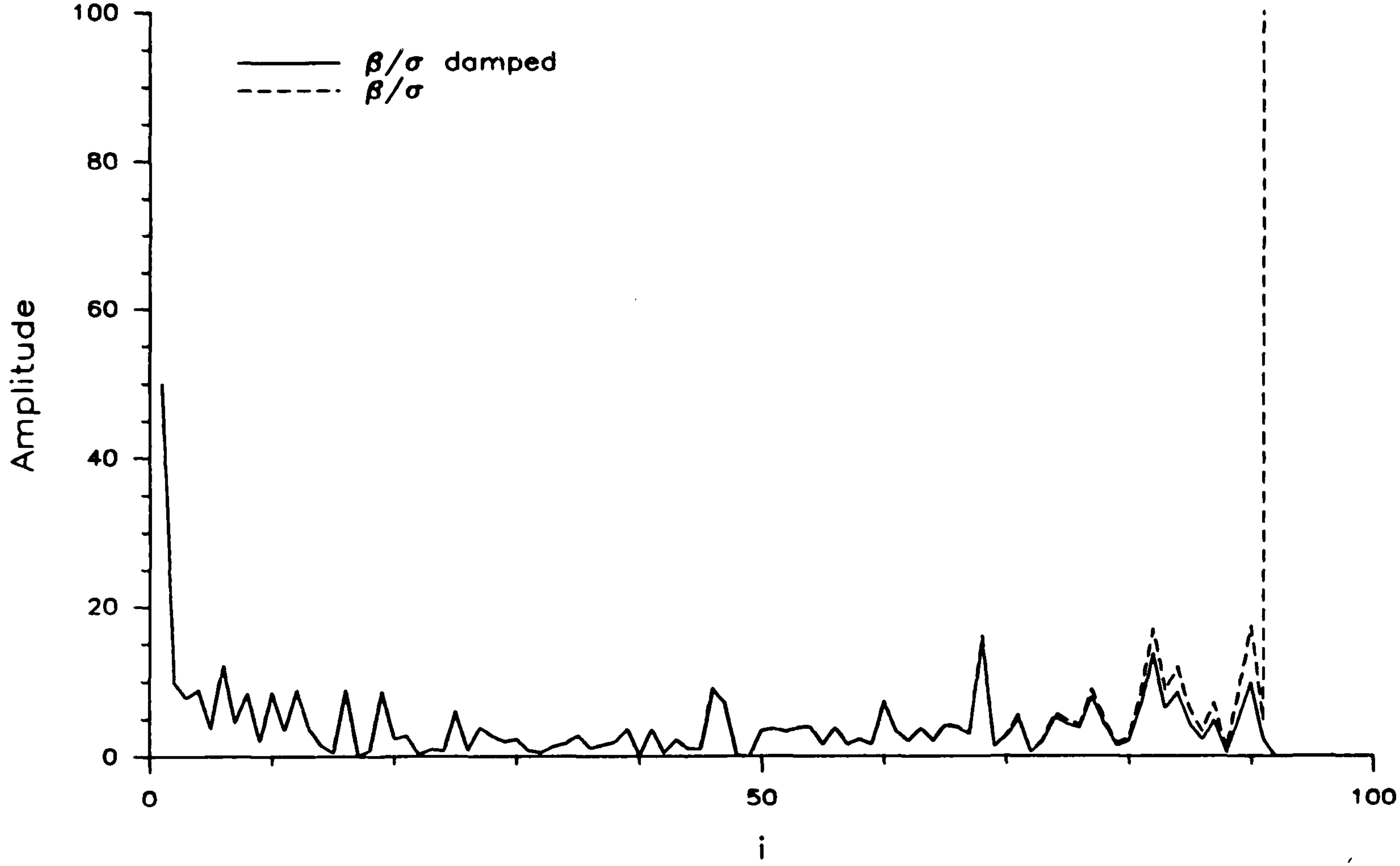


Fig. 4.3. The ratio shown in Fig. 4.2 when the singular values are damped by a regularization parameter.

for imposing the constraints, as mentioned in Chapter III, is to limit the space of solutions and therefore produce some kind of regularizing effect. Therefore, we are not necessarily interested in a solution which lies on the constraint surface (i.e., has elements with zero values). For this reason another method of applying constraints is needed.

Fig. 4.4 is an illustration which suggests another method of applying constraints. This figure depicts a least squares solution in two dimensions (i.e., x has two components). In this figure, e_1 and e_2 are the usual euclidean coordinates, with the constraint surface being the boundaries of the first quadrant. u_1 and u_2 are the transformed coordinates given by the SVD which are guaranteed to be orthogonal. The components of the solution vector along these coordinates are given by x_1' and x_2'. This solution vector is labeled x_{LS} in the figure. The process of projecting the solution onto the constraint surface (e_1, in this case) is indicated by the dashed line.

The method we suggest for applying constraints is summarized in the following:

a) Search for a feasible component of the solution (x_2' here).

b) Search for another component of the solution which when added to the current solution does not move the solution outside of the constrained region. In the example of Fig. 4.4, no such component exists. If such a component exists, repeat this step, otherwise, go to c).

c) If no component was found in b), then we assume an error in the measurement vector, b, is the cause of no solution being found. This error is assumed to cause an innacuracy in the magnitude of the corresponding component. Therefore, choose the component with the least error (i.e. it takes the solution the least distance outside of the constrained region), and the magnitude of the component is reduced such that the solution

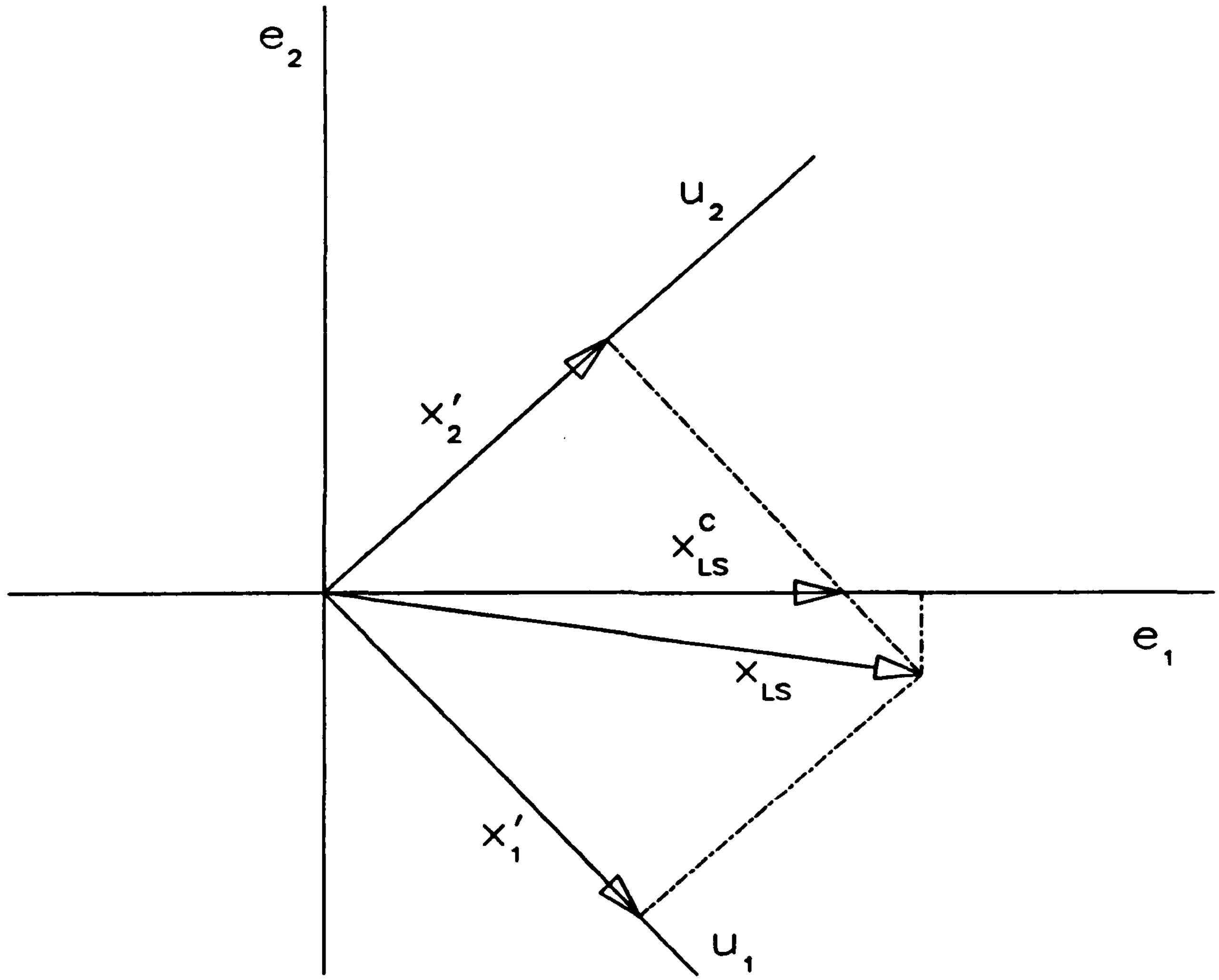


Fig. 4.4. Illustration of applying constraints when using the SVD algorithm for a two-dimensional example.

remains in the feasible region. The result of this process is x^c_{LS} in Fig. 4.4.

d) If all components have been used, then stop; otherwise, go back to b).

Note that as a result of using this method, the solution, x^c_{LS}, in Fig. 4.4 is closer to the u_2 axis than the solution obtained by projecting x_{LS} onto the constraint surface. This reflects the greater confidence we have in the x_2' component.

The assumption made above that the source of the error be restricted to the measurement vector may not be valid since numerical and modeling errors could cause the 'u' coordinates to be inaccurate. However, it would be more difficult to characterize and correct for these errors. Therefore, the method summarized above represents a good compromise between properly handling constraints and simplicity of implementation.

4.3.6 Conclusions

The SVD algorithm was shown to be a viable algorithm for the solution of the image reconstruction problem. Methods for regularizing the output of this algorithm and applying constraints were developed. This algorithm does suffer from the same shortcomings of any direct algorithm when applied to large sparse systems of equations. However, some of the methods developed in this section will also be applicable to the iterative algorithms which will be given greater emphasis in reconstruction process.

4.4 Iterative Algorithms for Image Reconstructions

4.4.1 Introduction

Iterative algorithms are very attractive for image reconstruction problems featuring large, sparse, coefficient matrices. The advantages of iterative techniques in this case include the following.

a) The sparsity of the coefficient matrix can be used to reduce computer storage and time requirements. This is not always the case for direct inversion algorithms.

b) Since in most iterative schemes the coefficient matrix is only used to form matrix-vector products, the coefficient matrix can be stored out of core and brought into main memory one row at a time.

In addition to these advantages for large, sparse systems, iterative techniques have the following advantages for general reconstruction problems:

a) As mentioned in Chapter III, the selection of stopping criteria for the iterative process can be used as a method of regularization.

b) Iterative techniques are more amenable to the application of constraints than direct inversion methods. This is because constraints can be applied at each step of the iterative process.

We will be interested in two major iterative techniques in this section. The first operates by attempting to successively satisfy each equation in Ax=b, and the second seeks to minimize a functional which results in a solution to the problem.

4.4.2 Projection Methods

A. Basic ART Algorithm

The projection method described in this section was originally developed by Kaczmarz as a means of inverting a large system of equations [70]. It was rediscovered as a method of solving the

tomography problem and renamed the algebraic reconstruction technique (ART) in the early 70's (see, for example [71] for an early paper on the application of the ART algorithm to computed tomography). Since its first use, it has remained one of the most popular methods for solving the system of equations resulting from the discretized tomography problem. Perhaps the best way to understand the ART algorithm is to take a geometrical point of view in the style of Tanabe [64]. The following notation is needed:

1) a_i = i^{th} row of A matrix

2) β_i = i^{th} element of b vector

3) $\langle a_i, x\rangle = \sum_{i=1}^{n} \alpha_i \xi_i$ = inner product of a_i with x

4) $\|x\|^2 = \langle x, x\rangle$

where the α_i's are the elements of the a_i vector, and the ξ_i's are the elements of the x vector. In this way, Ax=b is seen to be a set of m hyperplanes in n-dimensional Euclidean space, where the equation of the i^{th} plane (which will be called H_i) is seen to be

$$H_i : \langle a_i, x\rangle = \beta_i . \qquad (4.18)$$

If the system of equations is consistent, these planes will intersect in a unique vector, x. The system of equations is almost never consistent in practice because of measurement noise and the error inherent in the discretization process. The ART algorithm can be summarized as follows:

1) Choose an initial image vector, x_0.

2) Project this vector onto the hyperplane $\langle a_1, x_1\rangle = \beta_1$, using the equation

$$x_1 = x_0 - \frac{\langle a_1, x_0 \rangle - \beta_1}{\|a_1\|^2} a_1 \tag{4-19}$$

3) Iteratively project the result of (4-19) onto successsive hyperplanes such that the k^{th} iterate has the form

$$x_k = x_{k-1} - \frac{\langle a_k, x_{k-1} \rangle - \beta_k}{\|a_k\|^2} a_k \tag{4-20}$$

where k is greater than 1 and less than m. The iteration described by (4-20) is repeated until the updated image vector, x_k, has achieved some level of convergence. Note that in applying (4-20) if k=m, then set k=1 and k-1 = m in the next iteration.

B. Applying Constraints

In the form described above, no constraint has been made on the image vector, i.e. some elements of x_k may become negative, which is not physically possible. An inequality constraint of the form

$$\xi_i \geq \delta, \quad \text{for } i = 1,n \tag{4-21}$$

may be applied by choosing the x_k which is closest to the one given by (4-20) and also satisfies (4-21). That is, if some ξ_i is less than δ at the k^{th} step, then project x_k back onto the hyperplane $\xi_i=\delta$. A typical set of ART iterations is pictorially shown in Fig. 4.5 for m=3, n=2, and with inconsistent data. The constraints are ξ_1, $\xi_2 \geq 0$.

C. Properties of the ART Algorithm

Some of the basic properties of the ART algorithm will be listed in this section. For proofs of these results, see [64]. These properties will be useful for understanding the convergence properties of the algorithm. The first property is that each iteration can be written as the vector sum of an orthogonal projection of the latest iterate and the k^{th} row of A scaled by the k^{th} measurement. That is, (4-20) can be expressed as

$$x_k = P_k x_{k-1} + \frac{\beta_k}{\|a_k\|^2} a_k, \tag{4-22}$$

where P_k is an orthogonal projection matrix.

If we consider a set of iterations over all m rows of A then we can express these iterations as

$$x_m = Qx_0 + Rb, \tag{4-23}$$

where

$$Q = P_1 P_2 \ldots P_m, \tag{4-24}$$

and

$$Rb = \sum_{i=1}^{m} \frac{\beta_i}{\|a_i\|^2} Q_{i-1} a_i . \tag{4-25}$$

Equation (4-23) is important since if x_0 is contained in the kernel of A, then

$$Qx_0 = x_0 . \tag{4-26}$$

Therefore, if an initial vector (x_0) is chosen with a component in the null space of A, this component will appear in the final solution.

The final property which will be discussed is that (4-20) converges to the generalized inverse solution plus a projection of x_0 onto the null space of A. That is,

$$\lim_{k\to\infty} x_{k+m} = A^{+}b + P_{|\mathcal{N}(A)}x, \tag{4-27}$$

where $P_{|\mathcal{N}(A)}$ is the projection operator onto the null space of A. Note that if the equations are inconsistent, the convergence will be cyclical such that no matter how large k is, x_{k+1} will differ from x_k (see Fig. 4.5). However, for large k, x_{k+m} is close to x_k.

D. Underrelaxation for the ART Algorithm

The ART algorithm as described above has one problem in that for inconsistent equations the algorithm never converges to a solution. This behavior can be seen in Fig. 4.5. A means of avoiding this behavior is to use a relaxation parameter λ in (4-20) [72]. By including this parameter, (4-20) becomes

$$x_k = x_{k-1} - \lambda \frac{\langle a_k, x_{k-1}\rangle - \beta_k}{\|a_k\|^2} a_k. \tag{4-28}$$

For this relaxation method, the authors in [72] show that

$$\lim_{\lambda\to 0} x^*(\lambda) = A^{+}b + P_{|\mathcal{N}(A)}x_0, \tag{4-29}$$

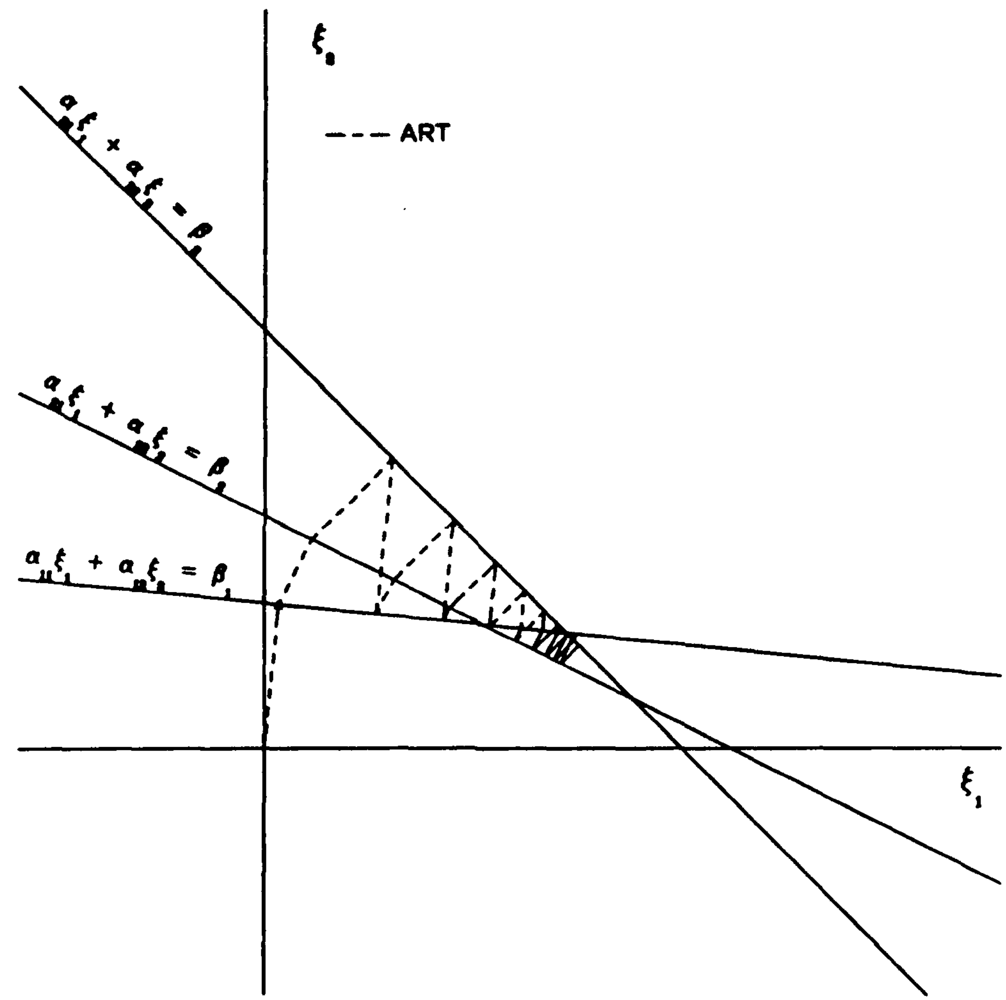


Fig. 4.5. Illustration of ART iterations for an inconsistent set of equations.

where

$$x^*(\lambda) = \lim_{k\to\infty} x_k(\lambda). \tag{4-30}$$

Note that this is better than (4-27) in that for large k, $x_{k+1}(\lambda)$ is close to $x_k(\lambda)$. Therefore, we can avoid cyclic convergence by choosing λ sufficiently small. However, for $0<\lambda<1$, (4-28) represents an incomplete projection onto the hyperplane H_k, of (4-18). That is, x_{k+1} does not reach the k^{th} hyperplane (hence the name underrelaxation). This implies that a small λ will mean slow convergence.

E. Other Projection Methods

Other methods have been developed (see [73] for an overview) which are similar to the ART algorithm in that at each iteration a projection is taken onto a convex set. For the ART algorithm, the convex set is the half-space defined by H_i, of (4-18). These different methods are obtained by giving different interpretations to the desired solution. For example, if the goal is to satisfy the i^{th} equation in $Ax=b$ within some bounds, then we are led to searching for solution in the intersection of m (number of equations) hyperslabs. These hyperslabs will be defined by the equations and the (error) bounds. This method is referred to as 'ART4' in [73]. In general, these other methods will converge to a solution in fewer steps than the ART algorithm, but will also require more computations per step. Instead of investigating these methods here, we wish to develop an iterative method which has some advantages over projection methods in the next section.

4.4.3 Gradient Methods

A. Introduction

In this section we consider solving an image equation of the form

$$Qx = b \tag{4-31}$$

where Q is an n×n symmetric positive definite matrix. Even though the matrix A will not be of this form in general, this assumption on Q will be made, but will be relaxed later. Also note that if the matrix A has full rank, then Q can be generated from the normal equations, (4-3).

B. Method of Steepest Descent

This method is a good introduction to the conjugate gradient algorithm, so it is discussed here. Solving (4-31) for the vector x is equivalent to minimizing the functional

$$f(x) = \tfrac{1}{2}x^T Qx - x^T b \tag{4-32}$$

over $x \in R^n$. This can be seen by taking the gradient of f(x) with respect to x, and setting it equal to the zero vector. We seek an iterative method such that

$$f(x_{k-1}+\alpha_k p_k) < f(x_{k-1}) \tag{4-33}$$

at each iteration. The p vectors are known as the search direction vectors, and α_k is the magnitude of the search. One choice for p_k is the negative gradient of $f(x_{k-1})$, giving the steepest descent direction for minimizing $f(x_{k-1})$. In this case p_k takes the form

$$p_k = b - Qx_k \qquad (4\text{-}34)$$

which is also known as the negative of the residual vector. The α_k which minimizes $f(x_{k-1}+\alpha_k p_k)$ can be found by expanding this functional as in (4-32). This α_k is found to be

$$\alpha_k = \frac{\langle p_k , p_k \rangle}{\langle Qp_k , p_k \rangle} \qquad (4\text{-}35)$$

If Q were allowed to be other than positive definite, as required, the denominator in (4-35) could become zero. The convergence of x_k to a solution vector, $\hat{x}$, can be quite slow if the ratio of the largest eigenvalue of Q to its smallest eigenvalue is large [61]. For this reason the method of steepest descent is not practical for solving (4-31). However, the conjugate gradient algorithm is able to achieve accelerated convergence even in this ill-conditioned case.

C. *Method of Conjugate Gradients*

Again, we wish to solve (4-31) by minimizing the functional given in (4-32). If we define p_1 through p_n as n, n×1 independent column vectors, then the span of these vectors will define an n-dimensional subspace

$$V = \text{span}\{p_1, \ldots , p_n\} \qquad (4\text{-}36)$$

The solution vector, $\hat{x}$, is obviously contained in V, so it can be written as the sum

$$\hat{x} = \alpha_1 p_1 + \ldots + \alpha_n p_n \tag{4-37}$$

The goal of the CG algorithm is at each iteration step k represent x_k on a k-dimensional subspace. Therefore, after n steps, if α_k is properly chosen, we will have found $\hat{x}$. This is the essential content of the expanding subspace theorem described in [74]. The CG algorithm also gives an efficient way to generate the vector space, V, such that the vectors, p_i, are Q-conjugate (i.e. $p_i^T A p_j = 0$ for all $i \neq j$). The algorithm that accomplishes this task was originally discovered by Hestenes and Stiefel [75]. An updated version is given here [61].

1) $x_0 = 0$

2) $r_0 = b$

3) $p_0 = 0$

For $k = 1, \ldots$

$$\begin{aligned} &\text{a)}\ \beta_k = \frac{\|r_{k-1}\|^2}{\|r_k\|^2} \\ &\text{b)}\ p_k = r_{k-1} + \beta_k p_{k-1} \\ &\text{c)}\ \alpha_k = \frac{\|r_{k-1}\|^2}{\langle p_k, Qp_k \rangle} \\ &\text{d)}\ x_k = x_{k-1} + \alpha_k p_k \\ &\text{e)}\ r_k = r_{k-1} - \alpha_k Q p_k . \end{aligned} \tag{4-38}$$

The first three steps above are initialization. In the iteration loop, steps a and b insure that the direction vectors, p_k's, are Q-orthogonal, while steps c and d minimize the functional in (4-32) by choosing x_k as the sum of the previous solution vector and a scalar multiple of the present direction vector. Step e in the iteration loop

is an efficient way of determining the residual. Note that in the algorithm above, no stopping criteria was specified for terminating the iterations, and no method of constraining the solution image was given. These two concerns will be given consideration in the remainder of this section.

D. Stopping Criteria

In any iterative algorithm it is important to determine at what point we should terminate the iteration process and output a solution. Not only is the determination of this stopping point necessary since it is inefficient to make additional unwanted computations, but as mentioned in Chapter III, stopping the iteration process can give a regularizing effect to the solution. Indeed, if the iteration process is continued too long, then a divergence from a good solution may occur.

The logical choice for deciding on a stopping point would be to proceed along the same lines as in the determination of a truncation point for the SVD algorithm. For the SVD algorithm the truncation decision was based on finding the point at which the error dominated in the (transformed) measurement vector. This was equivalent to ordering the singular values of the matrix A, and then finding the number of singular values to be used in the solution. Unfortunately, each step of the CG algorithm does not correspond to adding another coordinate (i.e., singular vector), in the SVD algorithm. Therefore, we cannot set the number of iterations equal to the truncation number found previously in the SVD procedure. If, however, at the k^{th} step of an iterative algorithm the solution, x_k, were to somehow depend on the singular values σ for $i<j$ (where j is to be determined), then we could use the results of the SVD analysis to choose a stopping point. Fortunately, this behavior has been observed in the CG algorithm [76], but the algorithm as presented in (4-38) does not include any mechanism

for observing and acting upon it. The LSQR algorithm presented in [76], which is analytically equivalent to the CG algorithm, does provide a mechanism for detecting the point at which 'j' singular values have been 'used' in the solution. The algorithm is outlined in Appendix C. In summary, the following steps should be performed for stopping the CG algorithm:

1) The singular value decomposition can be used to find the singular values and vectors of the matrix A.
2) The procedure for selecting which singular values should be used in the solution is given in the discussion on truncating the SVD algorithm.
3) Once this point has been determined to find a solution, use the algorithm in Appendix C, which bases its stopping procedure on the singular values, to find a solution.

E. Applying Constraints

As mentioned in Chapter III, the application of constraints on a solution vector (image) is of great importance. One of the advantages of iterative algorithms is the ability to constrain the solution at each iteration. In this way, errors will be corrected at each step.

The first method when considering the application of constraints (in the CG algorithm) is to project the solution onto the constraint surface after each iteration. As noted in [77], this procedure will not in general result in the generation of n linearly independent direction vectors. Therefore even if the least squares solution lies in the constrained region, it may not be found by this method. A solution to this problem is to restart the algorithm whenever the iterates have converged. The resulting algorithm is given by [77]

$$x_0 = 0$$

$$r_0 = 0$$

For $k = 0, \ldots, N$

1) Perform a CG iteration (4-39)

2) Apply constraints to x_k

3) If $\mathrm{ABS}(\|r_k\| - \|r_{k-1}\|) < \varepsilon$

restart by setting $p_k = r_k$,

where the CG iteration mentioned above is given by (4-38) or by the one in Appendix C, and ABS(∘) is the absolute value operator. Restarting causes the next search direction to be a steepest descent direction. See, for example, [78] or [79] for the convergence properties of restarted CG methods. Step 3 in the iteration loop above checks for convergence by observing the behavior of the residual vectors from one iteration to the next. This method has the following shortcomings:

1) At each step, the application of constraints may mean that the functional in (4-32) is not being minimized.
2) No criterion is given for choosing ε in (4-39), and ε may depend on the problem being solved.
3) The restart procedure may take the solution out of the feasible region.

For these reasons, an alternative method is desired for applying constraints.

The application of constraints can be based on the methods of constrained optimization discussed in [74]. The simplest is the method of feasible directions. This method could be adapted to the CG algorithm (4-38) by using the new update formula for x_k given by

$$x_k = x_{k-1} + \alpha_k' p_k \tag{4-40}$$

where α_k' is chosen such that its magnitude is the maximum value of α_k (given in 4-38) which leaves x_k inside the feasible region. As noted by the author [74], the feasible direction method can be subject to

jamming, which means that the solution does not change at each iteration. This jamming has been observed in the image reconstruction problem by the author. Of course, the algorithm could be restarted as above when jamming has occurred. Instead, we consider another method for applying constraints.

The gradient projection method was originally given by Rosen [80], as a method of maximizing or minimizing a functional subject to constraints. The method is also described in [74]. A summary of this technique follows for minimizing f(x) of (4-32).

1) Initialize x_0 in the feasible region (i.e., so that it satisfies all constraints).
2) Let $x_1 = x_0 + \alpha_1' p_1$, where α_1 and p_1 are chosen by the steepest descent method, and α_1' is the maximum value of α_1 such that x_1 is still in the feasible region.
3) If in step 2 a constraint surface were encountered, then define the working surface to be the (intersection of all) constraint surface(s).
4) Continue to minimize f(x) using $x_k = x_{k-1} + \alpha_{k-1}'$ over the working surface as in step 2.
5) If $f(x_k)$ is a minimum for x_k on the working surface, then determine, by moving in the direction of the negative gradient of $f(x_k)$, if $f(x_k)$ can be further minimized and x_{k+1} is still in the feasible region. If so, go back to step 4 with the corresponding constraint surface removed from the working surface; otherwise stop.

When this algorithm terminates, f(x) will be minimized over the feasible region. A more detailed description of the algorithm is given in Appendix D. The technique can be adapted to use the CG algorithm for minimizing f(x) and thereby avoid the slow convergence of steepest descent. Note that by moving in the direction of steepest descent in step 5, an implicit restart of the CG algorithm has occurred. Also

note that by minimizing f(x) over the working surface, the dimensionality of the problem has been reduced, and a computational advantage could be exploited.

An illustrative example of the procedure will be helpful. The example is shown in Fig. 4.6. The gradient projection method is used to solve the system of equations

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \tag{4-41}$$

subject to the constraints ξ_1 and $\xi_2 \geq 0$. The CG algorithm is used to generate the first direction vector. As can be seen from the figure, this direction is not feasible since it would result in ξ_2 being less than zero. Therefore, the working surface becomes the ξ_1 axis, and a minimum point of the functional (4-32) is found on this axis. At this point it is found that f(x) can be further minimized by moving in the direction of the negative gradient. Now, no working surface exists, and the CG algorithm finds the solution to (4-41).

4.5 A Comparison of the Algorithms

In this chapter, three algorithms have been presented for solving the matrix equation arising from the discretized tomography problem. Of these algorithms, the ART is the most well-known in the field of image reconstructions, and it has been thoroughly studied in the literature. The SVD algorithm is a standard algorithm for solving the least squares problem, although the method of handling constraints, which was developed here, may be new. The CG algorithm has been previously applied to the tomography problem [77],[81], but an adequate means of incorporating constraints was not given.

Before comparing these algorithms on some reconstruction

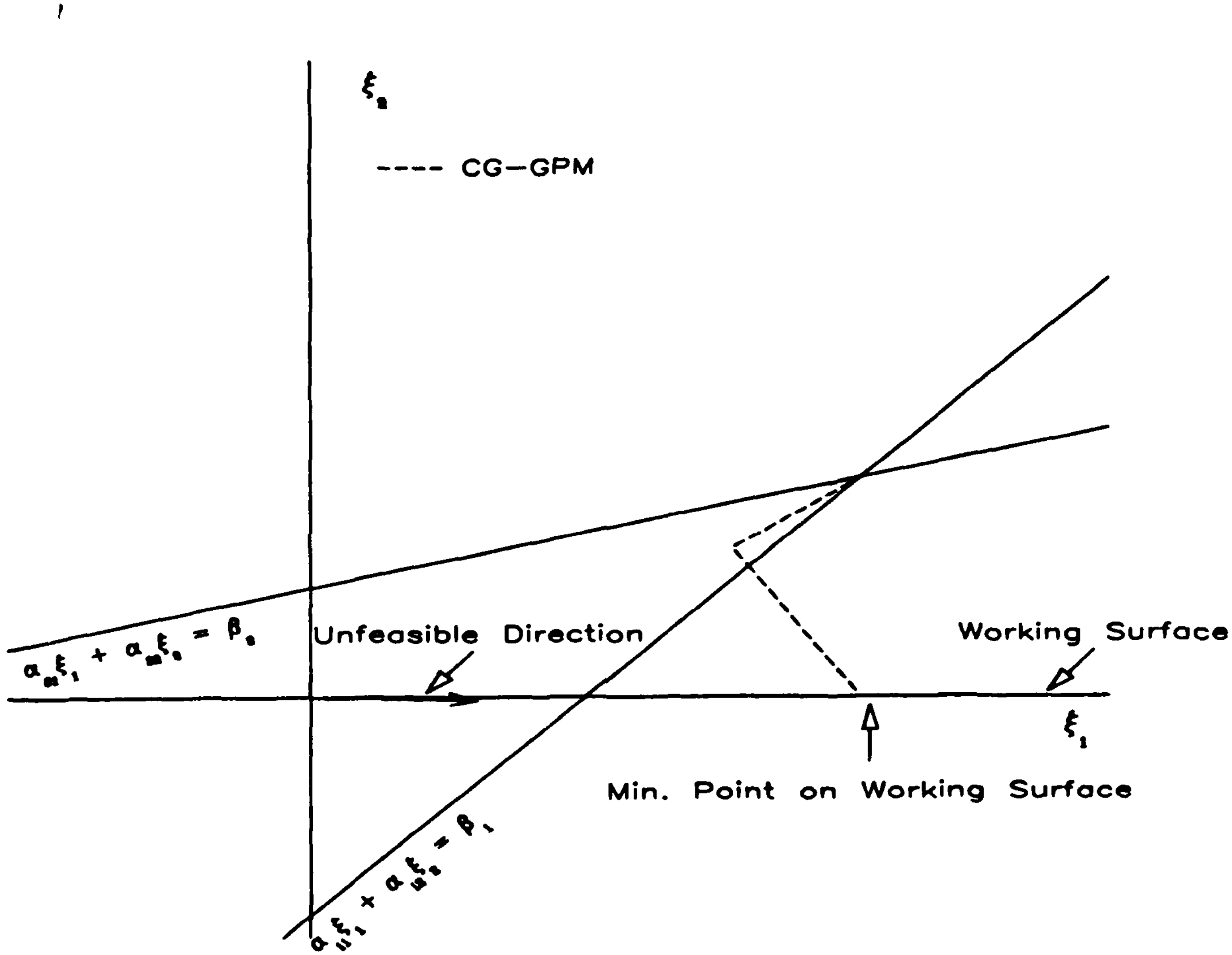


Fig. 4.6. Illustration of a sequence of steps in the gradient projection method.

problems, some graphical examples of their convergence behavior will be presented. Table 4.2 summarizes the various cases which will be examined. An explanation of the table is as follows: if the equations are consistent, then the lines defining the equations intersect in a unique point; if the LS solution is feasible then this solution satisfies the constraints; and finally if the equations are highly independent, then the lines do not have slopes which are nearly equal.

Table 4.2

Test examples for comparing the algorithms

Figure #	Consistent Equations?	LS Sol'n Feasible?	Highly Independent
4.7	Yes	Yes	Yes
4.8	Yes	No	Yes
4.9	Yes	Yes	No
4.10	No	Yes	Yes
4.11	No	No	Yes

In general, for the tomography problem, one would expect the equations to be inconsistent, the LS solution to be infeasible, and some of the equations to be almost dependent. So by examining the algorithms for the cases outlined in Table 4.2, some insight into their behavior for reconstruction problems may be obtained. For these figures, the algorithms will be denoted as follows:

a) SVD - The singular value decomposition with constraints applied as suggested in Section 4.2.

b) ART - The projection method applied to Ax=b (see 4-41). Unless otherwise noted, underrelaxation will not be used.

c) CG-GPM - The conjugate gradient algorithm applied to the gradient projection method. The CG-GPM operates on the normal equations, $A^TAx = A^Tb$.

The first example is shown in Fig. 4.7. This is the easiest problem to solve, and all three algorithms converge to the unique solution. The second example, which is shown in Fig. 4.8, has as

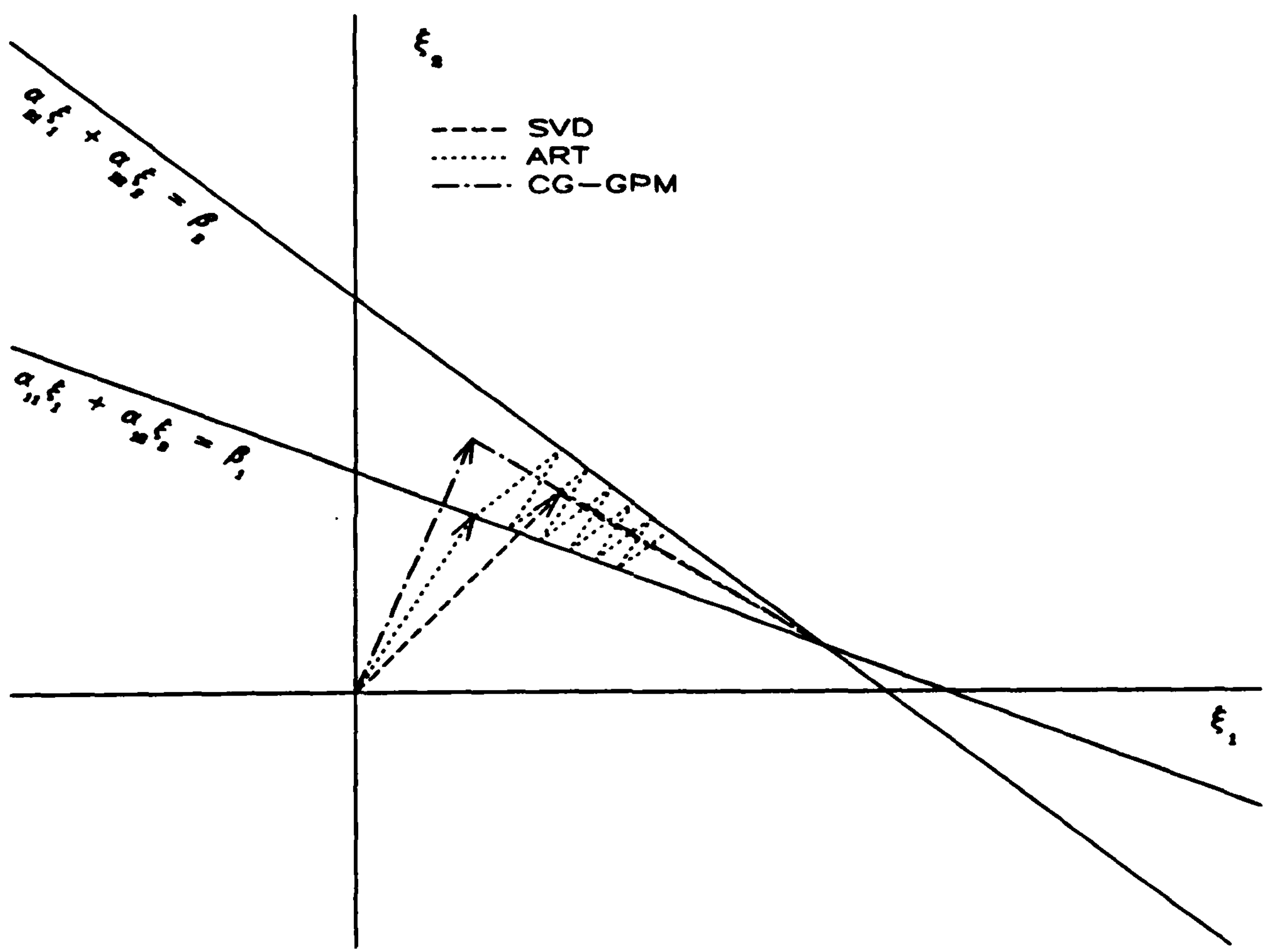


Fig. 4.7 Algorithm convergence behavior for a set of equations resulting in a consistent and feasible solution.

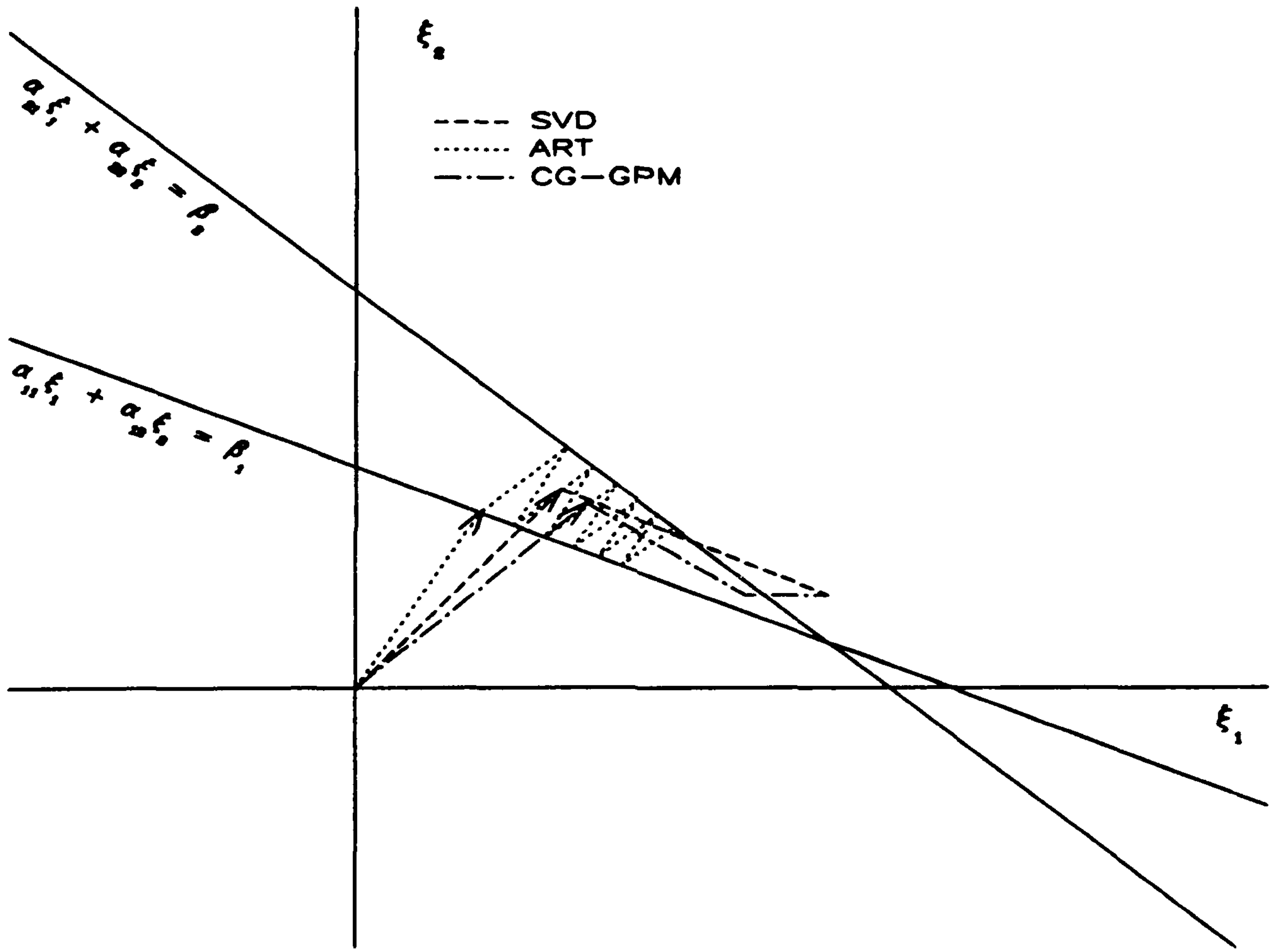


Fig. 4.8. Algorithm convergence behavior for a set of equations resulting in a consistent, but infeasible solution.

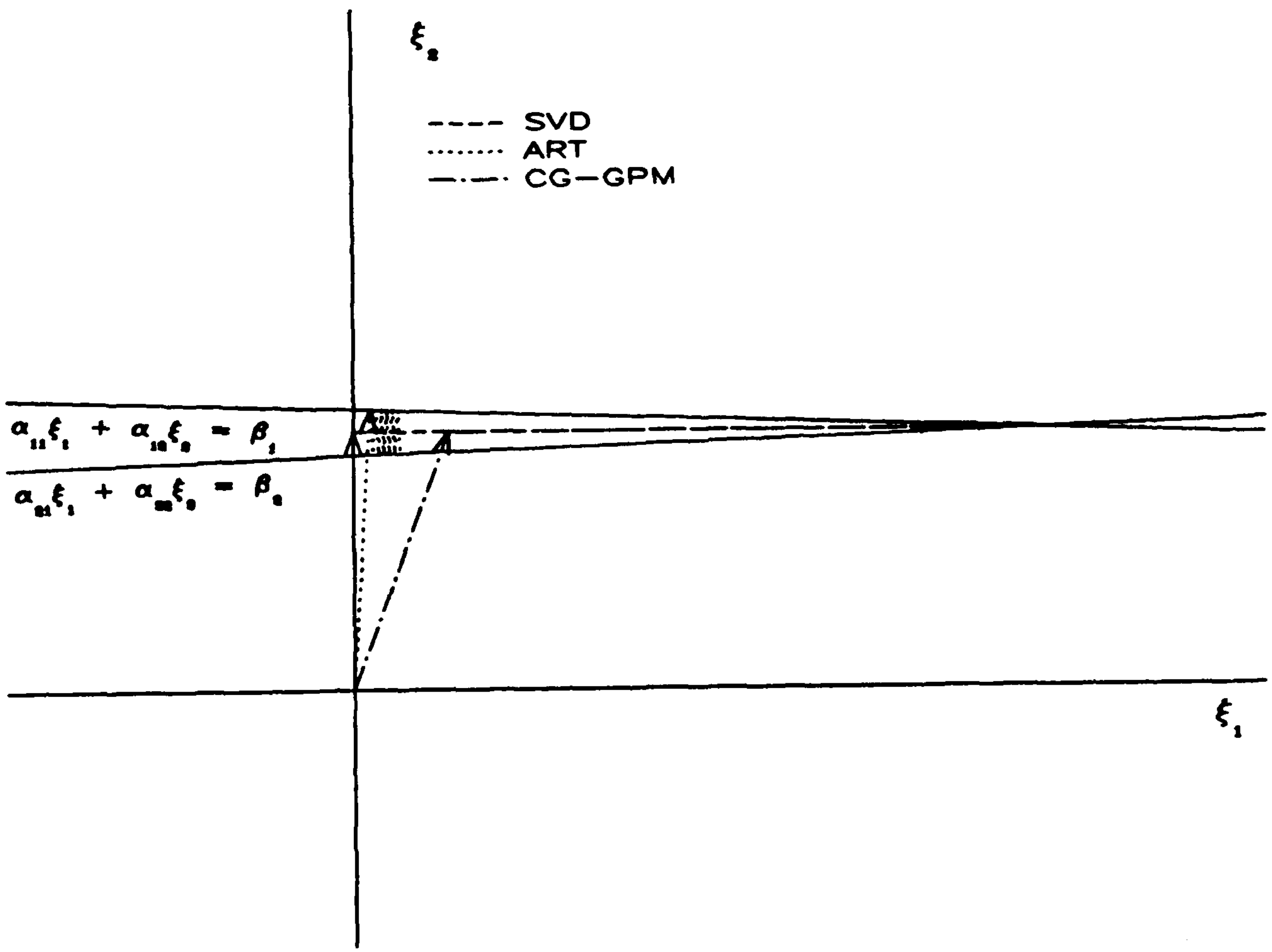


Fig. 4.9. Algorithm convergence behavior for a set of equations resulting in a consistent and feasible solution, the equations result in nearly colinear lines.

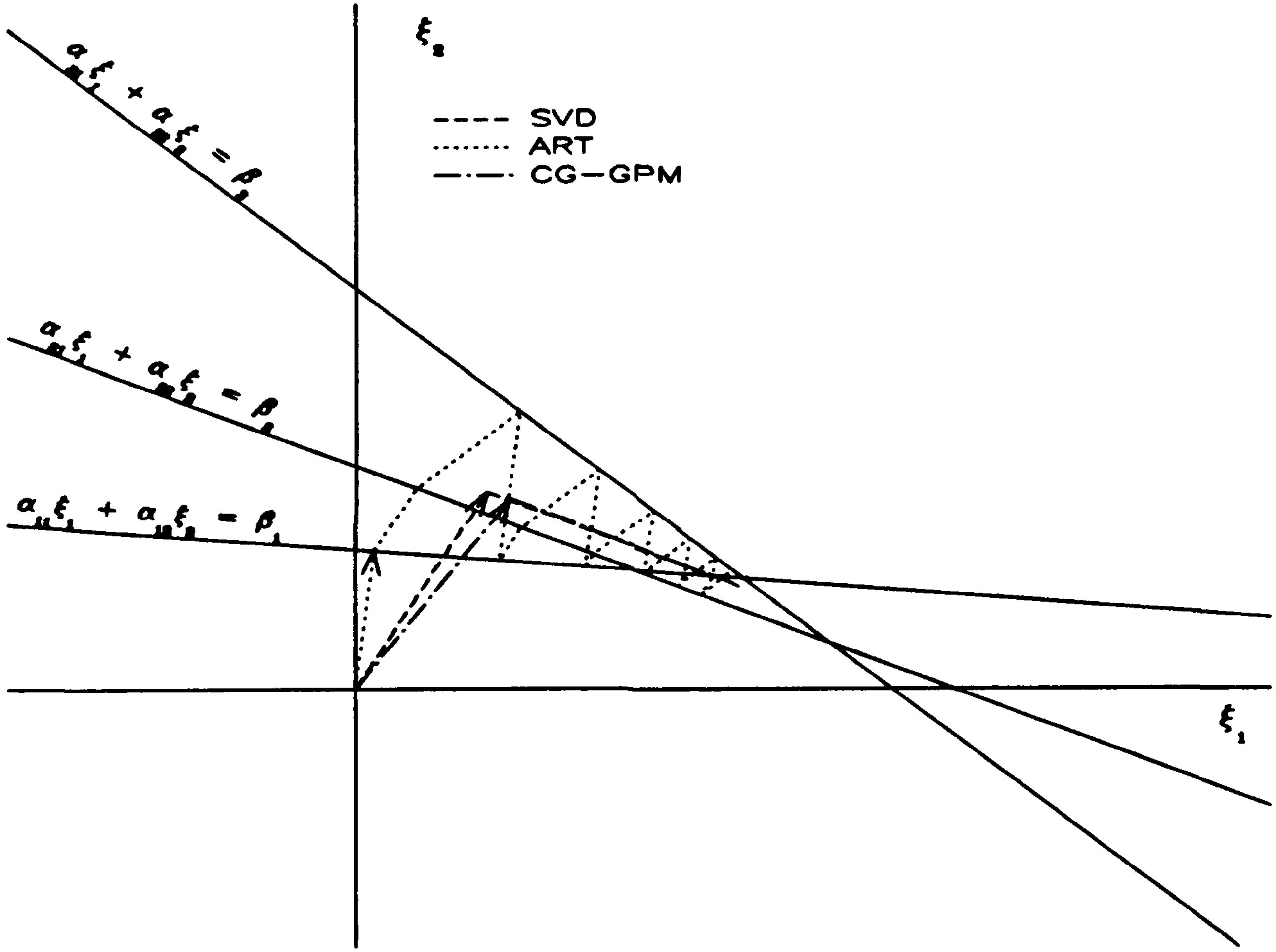


Fig. 4.10. Algorithm convergence behavior for a set of equations resulting in an inconsistent and feasible solution.

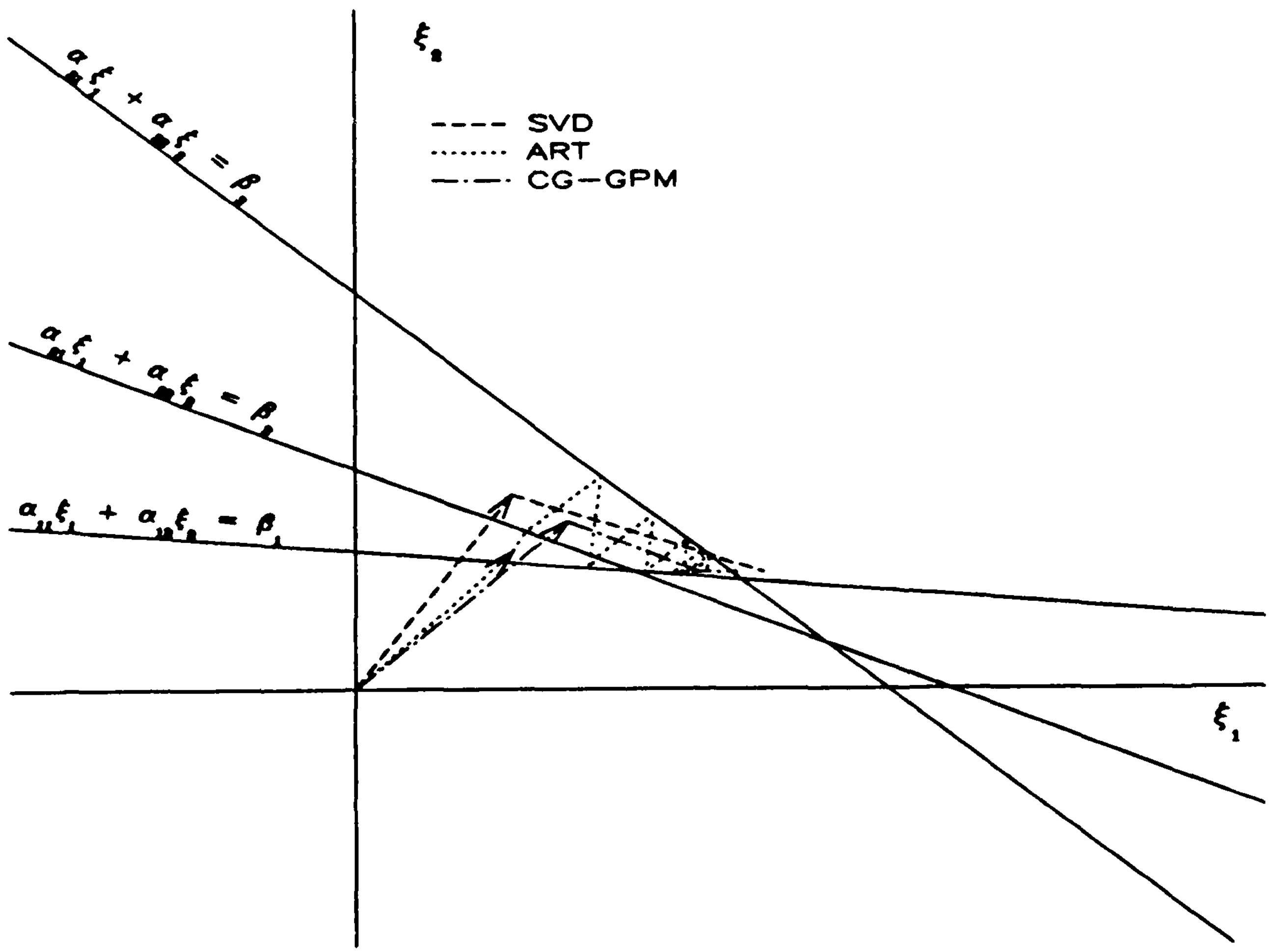


Fig. 4.11. Algorithm convergence behavior for a set of equations resulting in an inconsistent and infeasible solution.

constraints ξ_1 and $\xi_2 \geq 0.8$. As can be seen from the figure, both the SVD and CG-GPM converge to the same solution which is the point in the feasible region closest to the unconstrained solution. The ART algorithm never converges to a solution. The third example features a set of equation resulting in a pair of almost colinear lines (see Fig. 4.9). The SVD and CG-GPM algorithms show the same convergence to the solution (their graphs are in fact overlapping), while the ART algorithm shows slow progress towards the solution. The fourth example is for a set of three inconsistent equations (see Fig. 4.10). The CG-GPM converges to the minimum norm least squares solution given by the SVD. The basic ART algorithm would, of course, never arrive at the SVD solution. However, by using underrelaxation with a relaxation parameter of 0.8, convergence is obtained. The last example is for the same set of equations (see Fig. 4.11) as above with constraints ξ_1 and $\xi_2 \geq 1.0$. In this case, the CG-GPM algorithm is the only algorithm to converge to the point in the feasible region closest to the solution obtained above.

In summary, the GPM-CG algorithm appears to be the most flexible algorithm in that it arrived at the best solution in all five cases above. In addition, it is easiest to use since constraints are implicitly incorporated into the solution process, and no relaxation parameter needs to be chosen. In these examples, stopping criteria were not really an issue. Therefore the effect of choosing stopping criteria will have to be considered when actual reconstructions are performed.

CHAPTER V

IMAGE PROCESSING

5.1 Introduction

In Chapter IV, various method for reconstructing a cross sectional image from its projections were discussed. The images generated by these algorithms are often hard to interpret because of limitations in the measurement process and in the reconstruction model. For example, if diffraction effects are ignored by using the straight-ray model, then the reconstructed image will exhibit features (due to the diffraction phenomena) which are not present in the actual cross section.

Fig. 5.1 shows the reconstruction of a cross section of earth containing a circular conducting cylinder. The details of the process of obtaining this reconstruction will be discussed in Chapter VII. In this figure, the pixel attenuation values are given at the intersection of the horizontal and vertical grid lines and the higher attenuating regions are represented by the peaks in the figure. The peak in the center of the figure gives the location of the cylinder, but additional peaks are evident which are probably the result of diffraction effects (recall the ripple in the electromagnetic response of Fig. 2.29). The spreading of the cylinder location in the horizontal direction is a result of the limited view angle. It would be desireable to smooth out the peaks and valleys in Fig. 5.1 which do not represent the cylinder's location.

Fig. 5.2 is the result of applying a technique called selective smoothing [48] to the image in Fig. 5.1. Note that the unwanted artifacts in the image have been smoothed out with little effect to the peak which indicates the location of the cylinder. The goal in this chapter is to develop techniques, such as selective smoothing, which

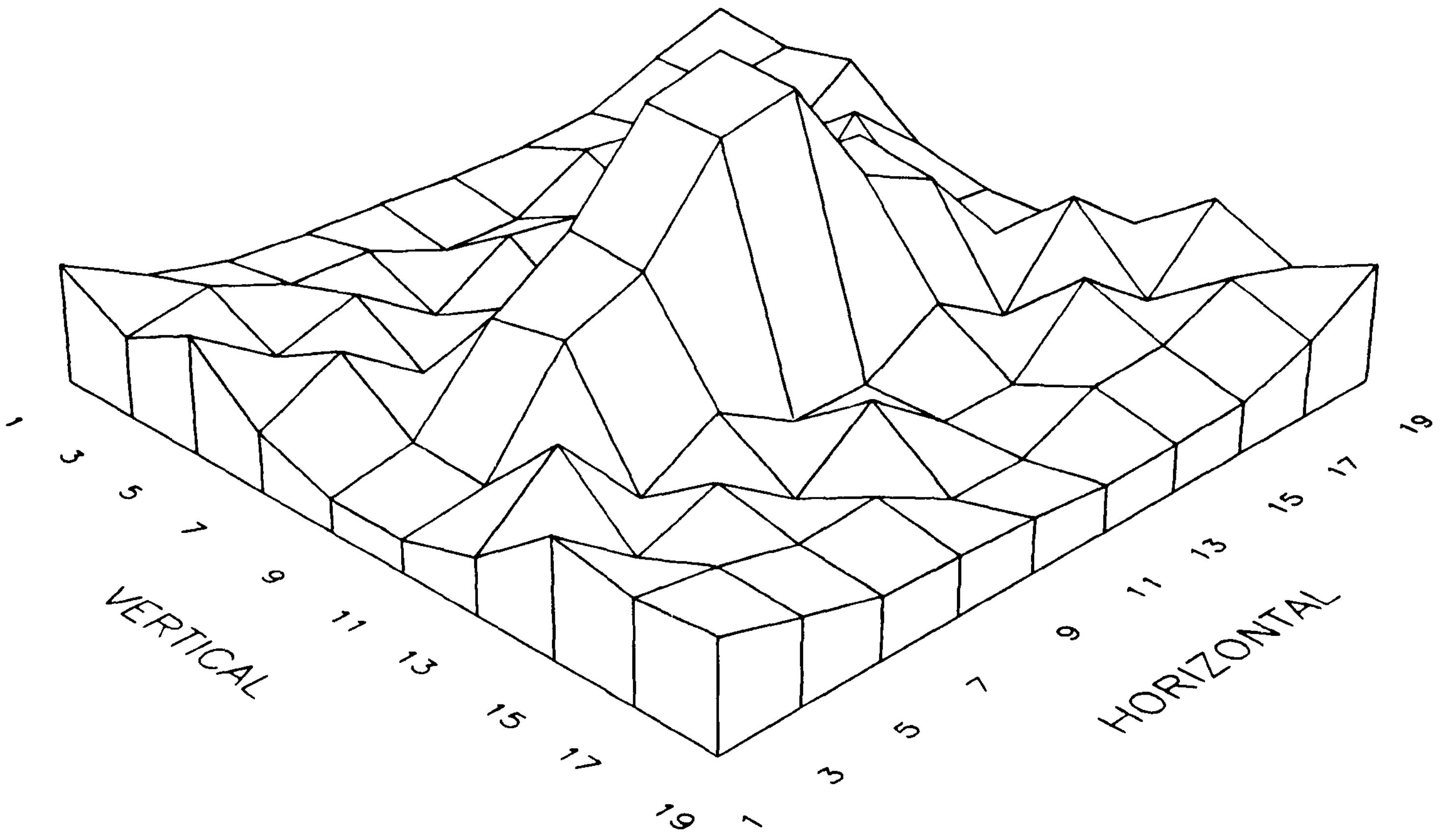


Fig. 5.1. Reconstructed cross sectional image of a homogeneous earth containing a single circular cylinder.

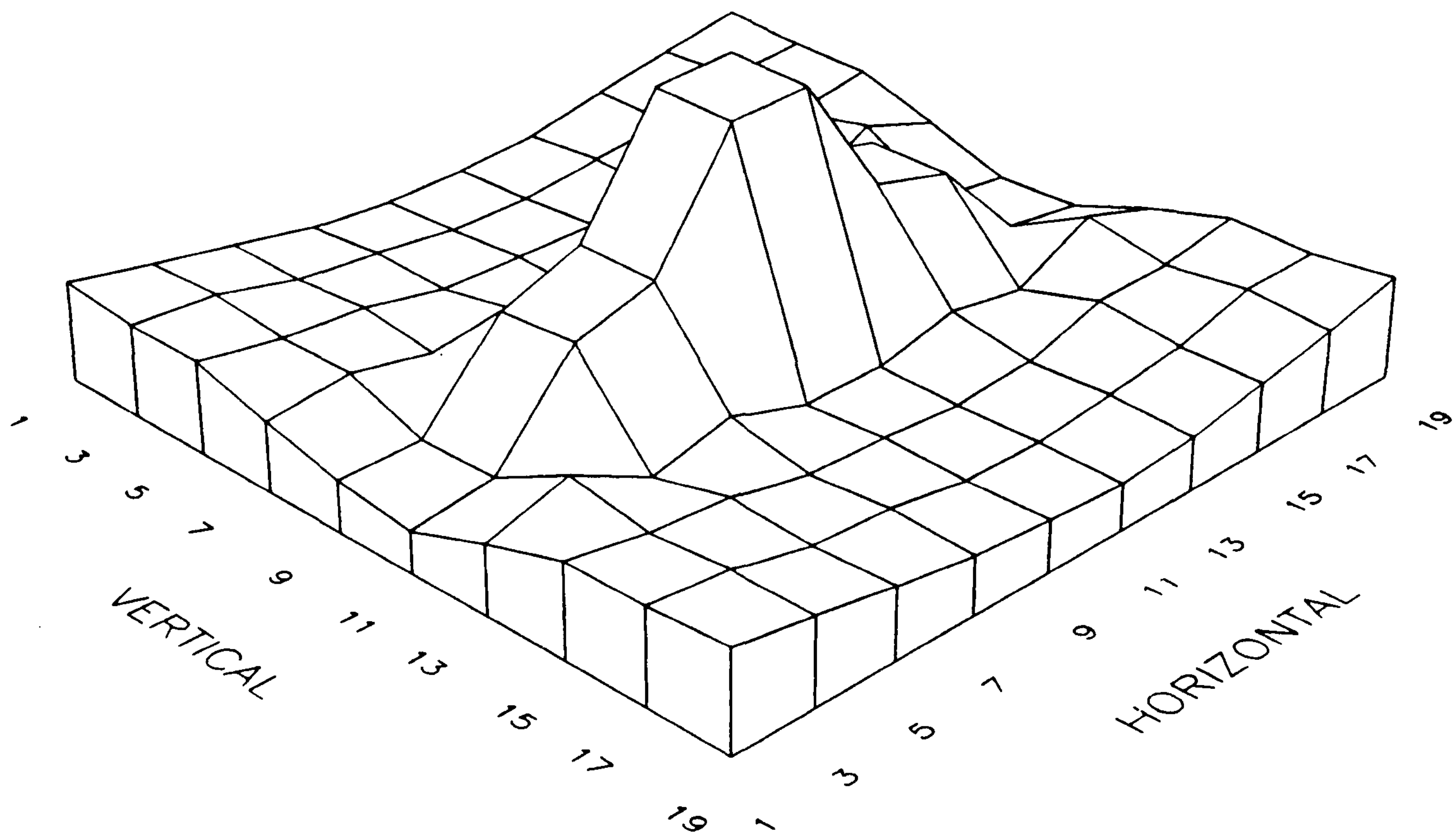


Fig. 5.2. The result of filtering Fig. 5.1 using selective smoothing.

will improve some feature(s) of a reconstructed image.

5.2 Image Filtering and Smoothing

5.2.1 Introduction

In the survey article on image enhancement [82], the authors list three methods of image enhancement that are used in the spatial domain. These are:

a) Spatial smoothing of regions using low-pass filters.

b) Gray level rescaling.

c) Edge enhancement using high pass filters.

Since our goal is to reduce unwanted artifacts in the image, we will consider only spatial smoothing of the reconstructed images.

5.2.2 Spatial Smoothing

Spatial smoothing techniques usually operate by passing a window over an image and then replacing the attenuation of the pixel at the center of the window by the weighted average of all the cells in the window. We will be interested only in nine cell windows (see Fig. 5.3), so that the attenuation at the center pixel is given by

$$\alpha'_{00} = \frac{1}{D} \sum_{i=-1}^{1} \sum_{j=-1}^{1} w_{ij}\, \alpha_{ij} \; , \tag{5-1}$$

where the indexing on α is indicated in Fig. 5.3, w_{ij} is the i-jth element of the weighting matrix, and D is the sum of all w_{ij}'s. If all w_{ij}'s are equal to 1.0, then (5-1) gives an equally weighted spatial smoothing filter. The equally weighted filter will tend to excessively blur the image. The following weighting matrix is able to achieve

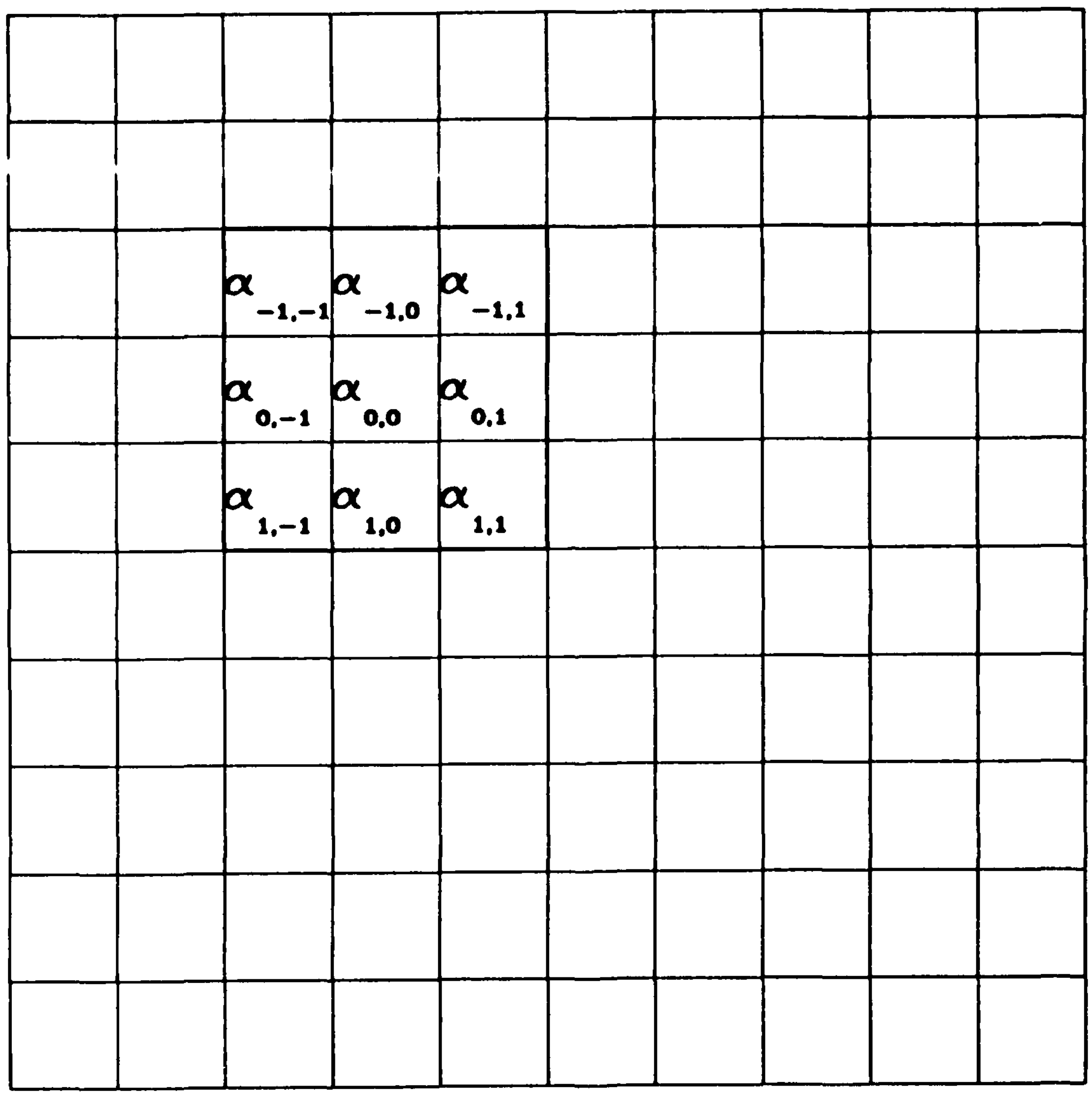


Fig. 5.3. Illustration showing a nine cell window for filtering an image.

smoothing without excessive blurring [82]

$$W = \begin{bmatrix} 0.25 & 0.50 & 0.25 \\ 0.50 & 1.00 & 0.50 \\ 0.25 & 0.50 & 0.25 \end{bmatrix}. \qquad (5\text{-}2)$$

Blurring has been reduced by giving the center cell larger weighting than the adjacent cells. Fig. 5.4 is the result of applying (5-1) using the weighting in (5-2) to the unfiltered image of Fig. 5.1. Note that the unwanted peaks and valleys have been smoothed out, but it is hard to distinguish the boundary of the cylinder.

A spatial filter which is able to smooth out the unwanted peaks and valleys without blurring the boundary between the cylinder and the background is the selective smoothing filter [48]. For this spatial filter the weighting matrix is given by

$$W = \begin{bmatrix} f(\alpha_{-1-1}) & f(\alpha_{-10}) & f(\alpha_{-11}) \\ f(\alpha_{0-1}) & 2 & f(\alpha_{01}) \\ f(\alpha_{1-1}) & f(\alpha_{10}) & f(\alpha_{11}) \end{bmatrix}, \qquad (5\text{-}3)$$

refer again to Fig. 5.3 for the definitions of the α's. The function $f(\circ)$ is nonlinear and is given by

$$f(\alpha_{ij}) = \begin{cases} 1 & \text{if } |\alpha_{ij}-\alpha_{00}| \leq T \\ 0 & \text{otherwise.} \end{cases} \qquad (5\text{-}4)$$

The threshold value, T, will determine the amount of smoothing to be performed. For example, if T is set to a large value, (5-3) reduces to a weighted average matrix. The idea behind using a threshold is that if a pixel has a much higher (or lower) attenuation value than surrounding pixels, its attenuation will remain unchanged. The threshold value should be some fraction of the difference between the highest and lowest value of attenuation in the unfiltered image. A value between 1/5 and 2/5 for this fraction was suggested in [66]. The

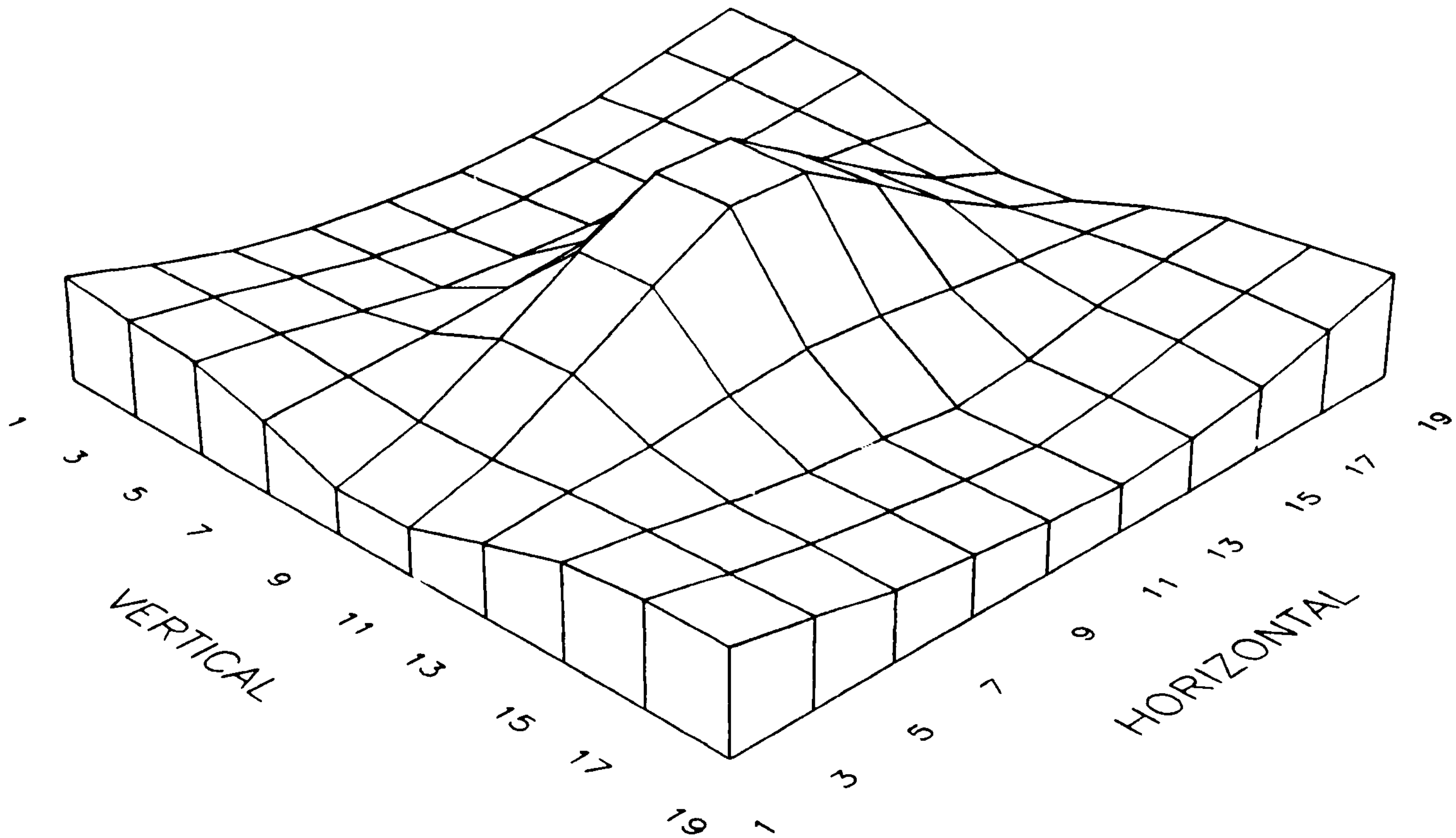


Fig. 5.4. The result of filtering Fig. 5.1 using a weighted average filter.

image in Fig. 5.2 is the result of using selective smoothing on Fig. 5.1 with the threshold value equal to 0.3 times the difference between the maximum and minimum attenuation values.

5.2.3 Spectral Filtering

We do not expect filtering in the frequency domain to be as effective as spatial filtering for reconstructed images. This is because the most effective spatial filter was nonlinear (i.e., 5-3), and therefore would not have a correspondingly simple frequency domain implementation.

Fig. 5.5 was obtained by taking the two-dimensional discrete Fourier transform (DFT) of the unfiltered image shown in Fig. 5.1. The zero frequency ('DC') point is in the center of the x-y plane. Fig. 5.6 is the DFT of the selective smoothed image shown in Fig. 5.2. Note that it is hard to determine the relationship between Figs. 5.6 and 5.5, except to note that regions of high frequency in both (i.e., simultaneously) the horizontal and vertical directions have been somehow attenuated. Fig. 5.7 is the DFT of the weighted average filtered image shown in Fig. 5.4, and again a relationship to Fig. 5.5 is hard to discern.

Fig. 5.8 shows how a low pass filter could be applied to the spectrum in 5.5. That is, set the high frequency components in the spectrum equal to zero. By taking the inverse DFT of this low-passed spectrum, the image in Fig. 5.9 is obtained. This image has been successfully smoothed, but the location of the cylinder is severely blurred.

5.3 Object Detection

Once a reconstructed image has been filtered, one often wants to perform some sort of processing of the image to separate regions (of

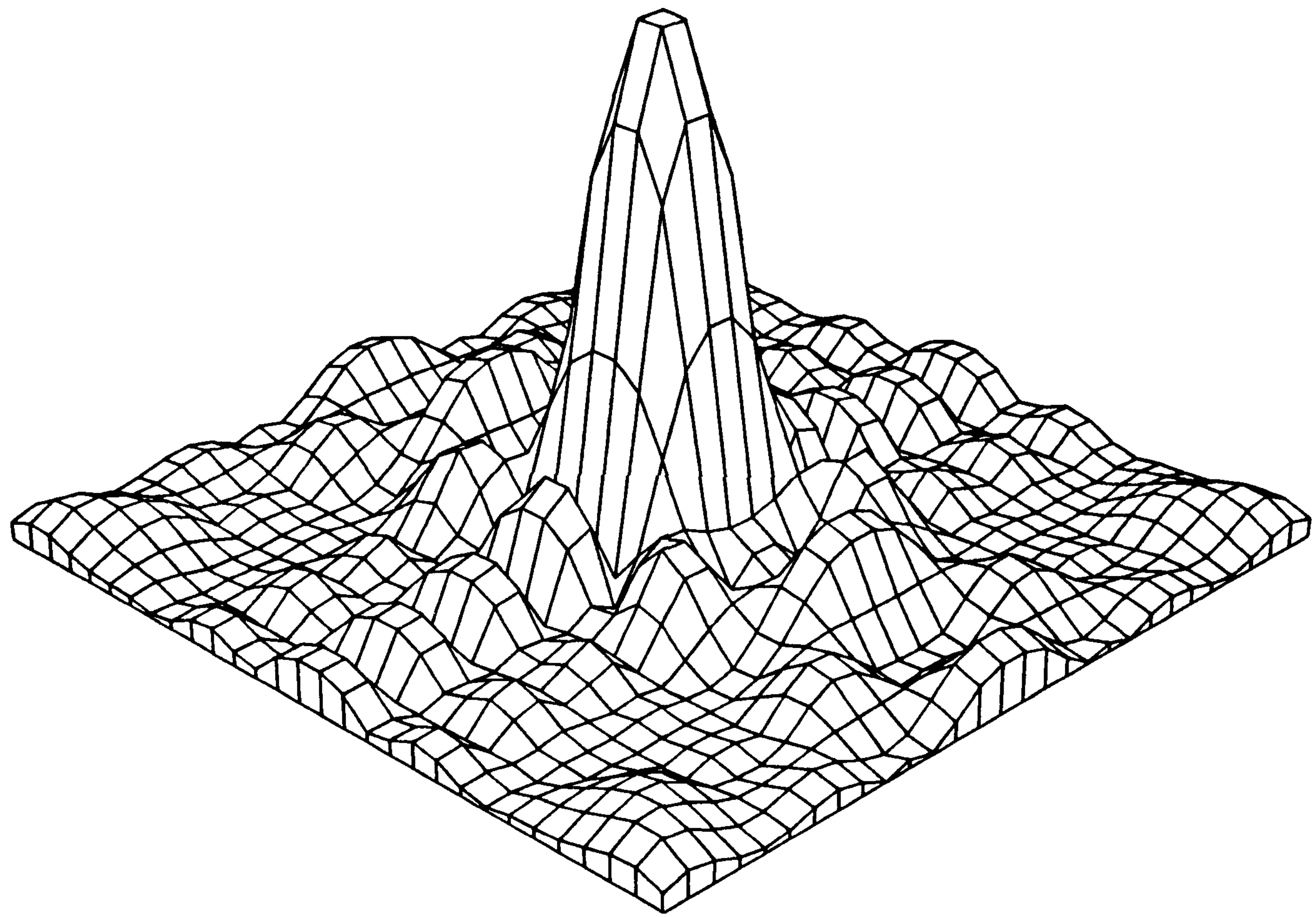


Fig. 5.5. Frequency domain representation of the image in Fig. 5.1.

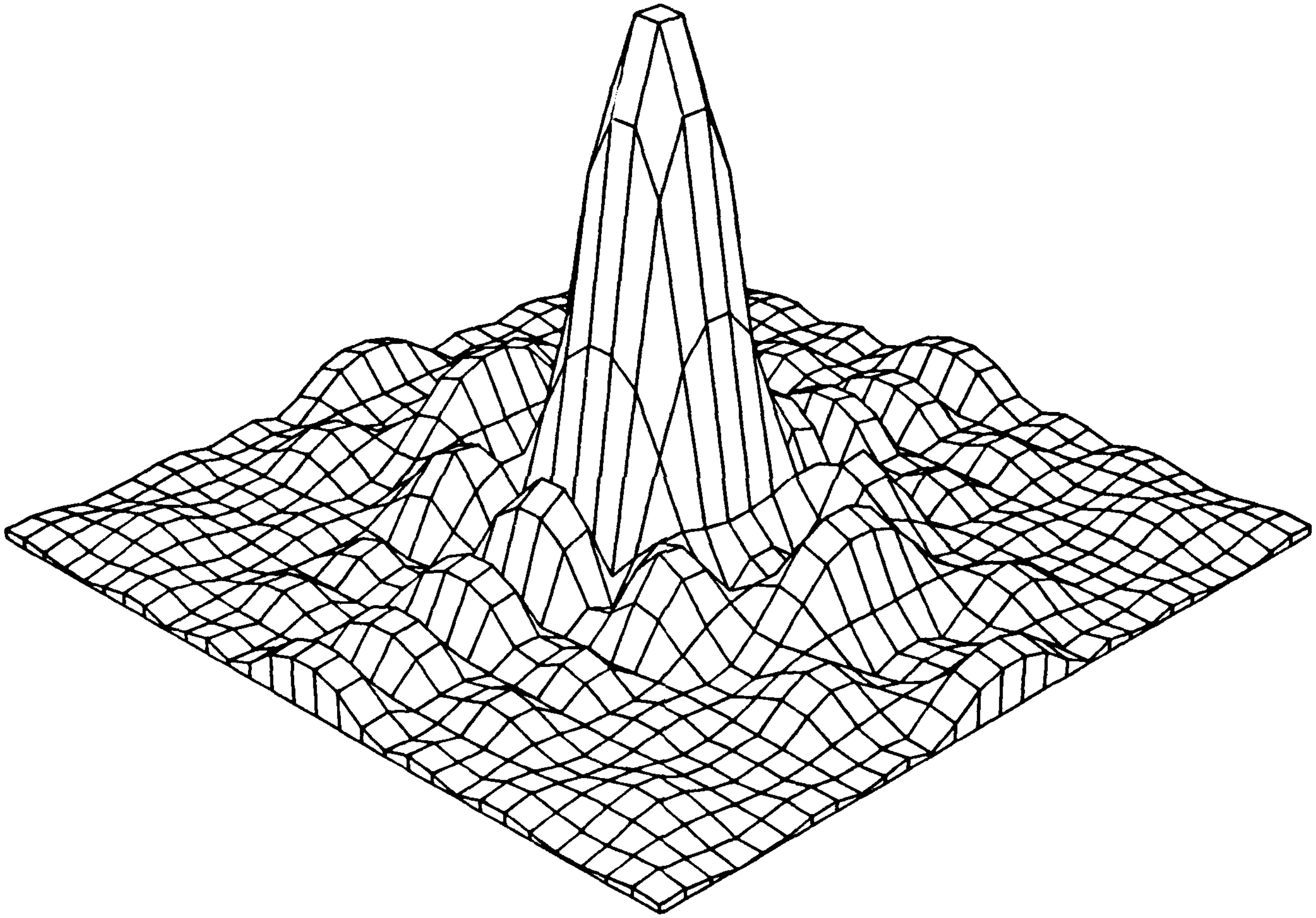


Fig. 5.6. Frequency domain representation of the image in Fig. 5.2.

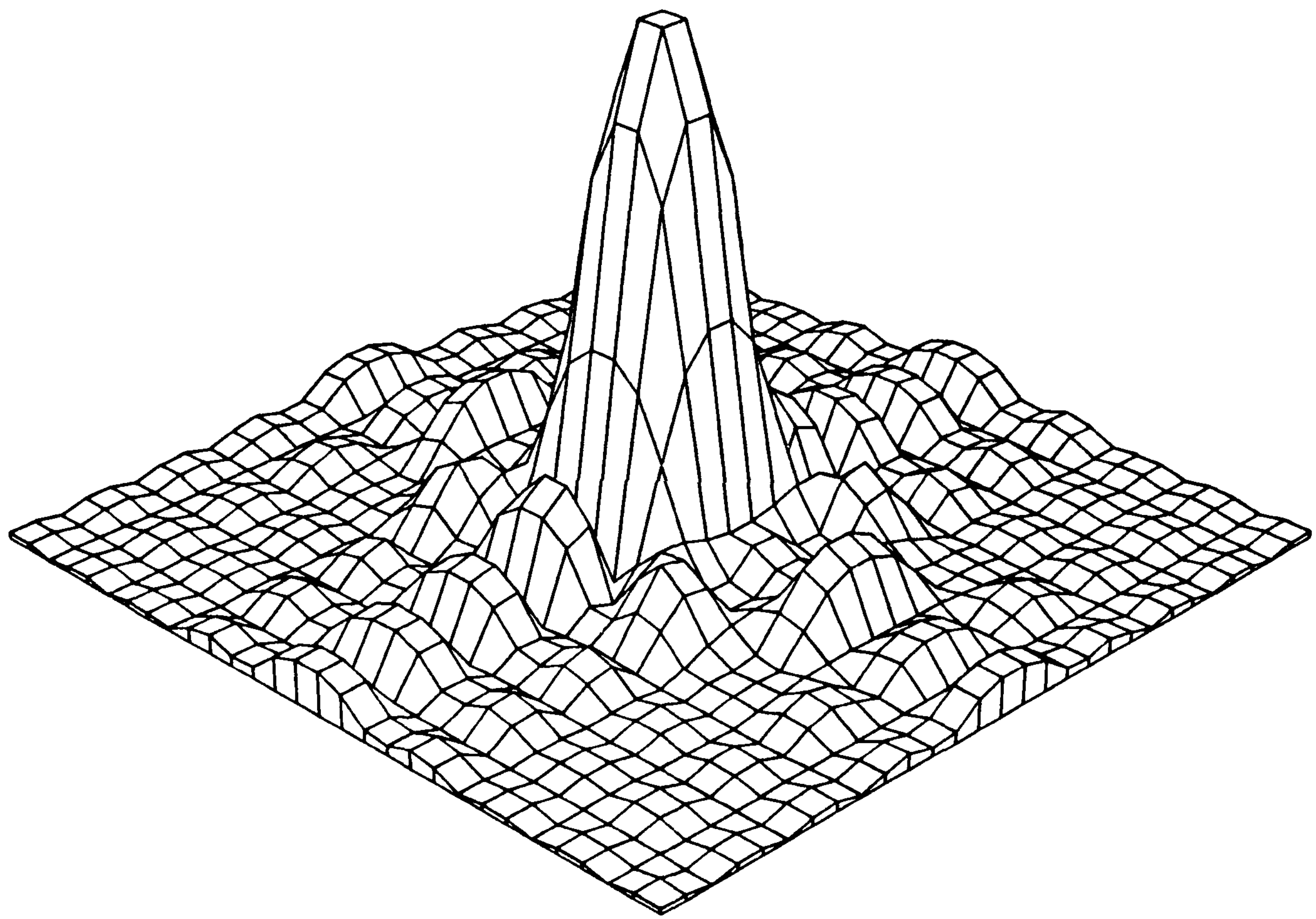


Fig. 5.7. Frequency domain representation of the image in Fig. 5.4.

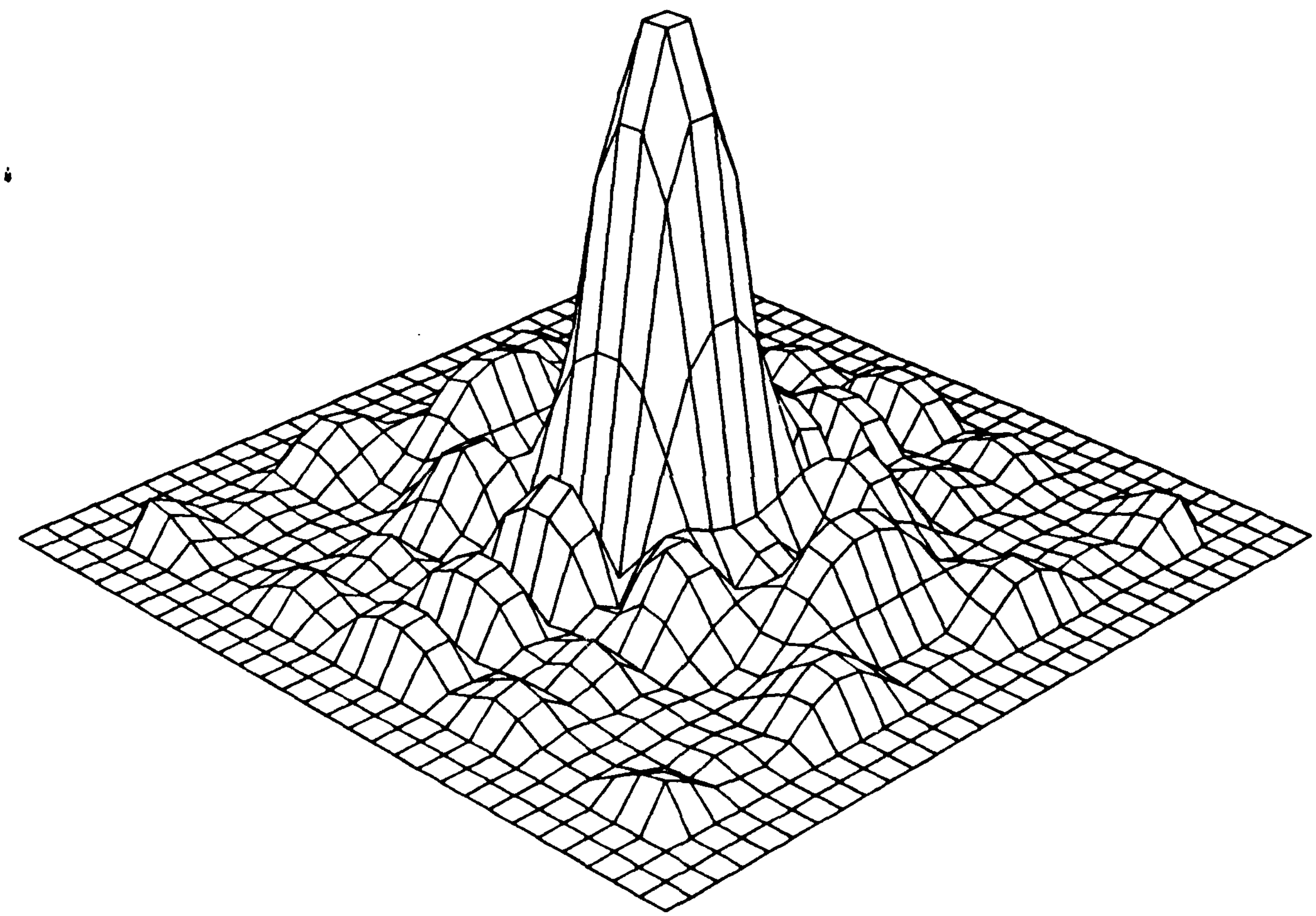


Fig. 5.8. The result of low pass filtering the spectrum shown in Fig. 5.5.

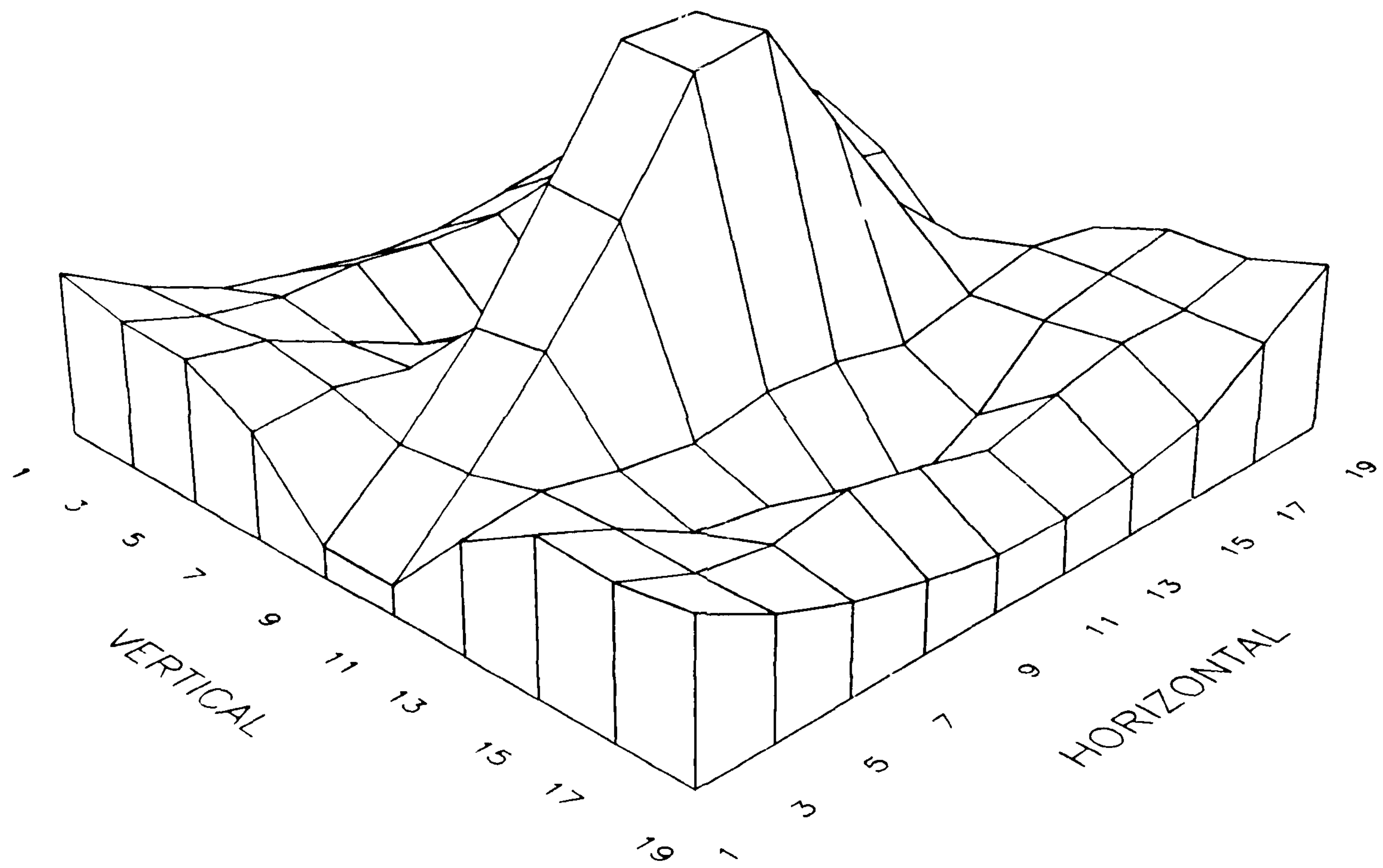


Fig. 5.9. The image obtained by inverse transforming the spectrum shown in Fig. 5.6.

pixels) having similar attenuation. This type of processing is called partitioning. In this way we may be able to discern anomalous regions from the background. Although this process can be accomplished by a human operator, a more objective means is desirable. For this reason we consider ways to partition the cells in a reconstructed image. The partitioning process involves placing the cells into groups (usually 2 or 3 groups) based on the attenuation value found during the reconstruction process.

Two types of partitioning will be used here. The two partitioning schemes are similar in that the groups are found by minimizing the distance between members of the group and the group average. The first scheme is called minimum variance partitioning (MVP) [83], and the distance function is given by

$$d_{MVP}(j) = \sum_{i \in group} (\alpha_i - \mu_j) \qquad (5\text{-}5)$$

where α_i is the attenuation coefficient of the i^{th} cell in the image and μ_j is the average value of the pixels in the j^{th} group. The second method we call minimum max partitioning (MMP), with a distance function defined by

$$d_{MMP}(j) = \max_{i \in group} |\alpha_i - \mu_j| \qquad (5\text{-}6)$$

That is, the distance is equal to the maximum difference between the group average and the members (i.e. pixels) of the group. This distance function should be used in cases where outliers (e.g., tunnels or oil deposits) may be present and need to be emphasized. Using this distance function would corresponds to a minimization problem in the L_∞ norm. As can be seen from comparing (5-5) and (5-6), the MVP will have more of an averaging effect than the MMP since all members of the group

are used in the distance function of (5-5). For either distance function, the quantity to be minimized will be given by

$$J_e = \sum_{j=1}^{K} d(j), \tag{5-7}$$

where, for example K=2 for a two-group partition. The quantity d(j) is given by (5-5) or (5-6). The objective is to find all partitionings of the pixels into K groups. Then, for each partitioning, calculate J_e in (5-7), and choose that partitioning which has the smallest J_e, thereby minimizing (5-7).

Fig. 5.10 shows the result of using the MVP on the smoothed image of Fig. 5.2. Note that pixels outside of the cylinder location have been included in the high attenuation group. Fig. 5.11, on the other hand, which uses the MMP, includes only adjacent pixels in the high attenuation region. Fig. 5.12 summarizes these results in two-dimensional displays, with higher attenuation regions represented by darker shading. The (a) part of the figure is for the MVP, while the (b) part of the figure is for the MMP. Also shown in both parts of the figure is the location (indicated by a circle) of the cylinder which was present when the data was simulated.

In summary, by taking the raw reconstructed image in Fig. 5.1, then filtering this image using selective smoothing, and finally performing the MMP, a very good indication (to the resolution of the pixels) of the actual location of the cylinder is achieved.

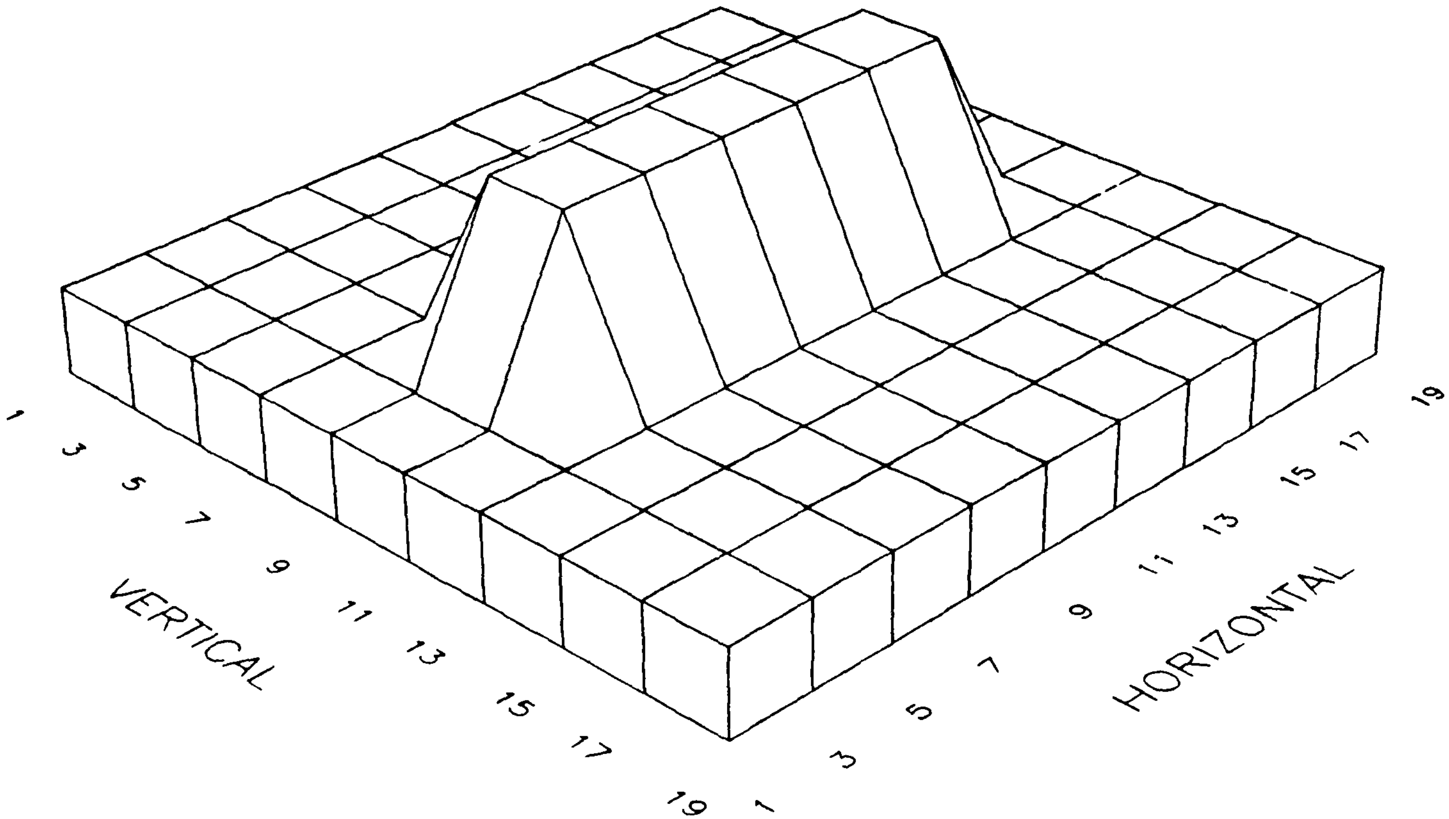


Fig. 5.10. Minimum variance partition of the image in Fig. 5.2.

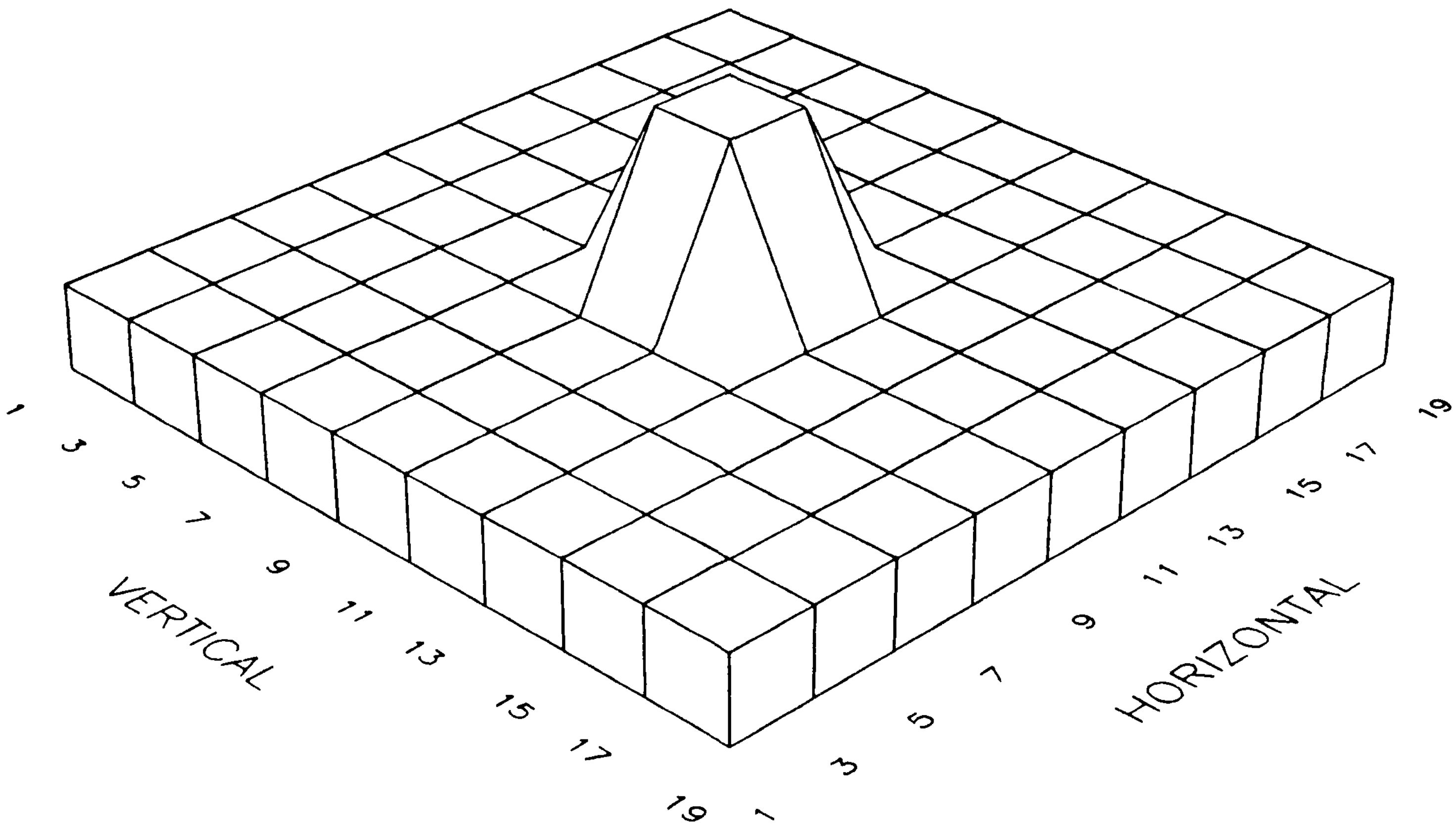
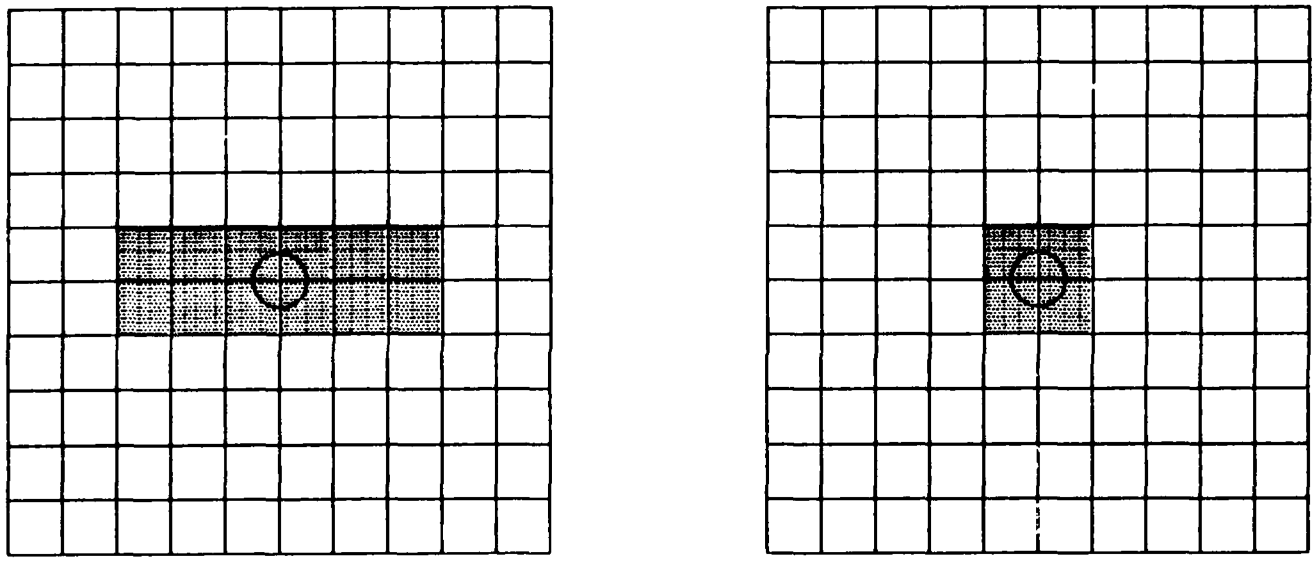


Fig. 5.11. Minimum max partition of the image in Fig. 5.2.



(a) (b)

Fig. 5.12. Summary cf partitionings: (a) MVP, (b) MMP

CHAPTER VI

COMBINING TIME-OF-FLIGHT AND CONTINUOUS WAVE MESUREMENTS

6.1 Introduction

Two methods we will be considering for taking measurements in a cross-hole system are:

a) Time-of-flight (TOF) measurement, in which the time it takes a radio frequency pulse to travel between transmitter and receiver is measured.

b) Continuous wave (CW) measurement, in which the amplitude ratio between a transmitted and received sinusoidal wave is measured.

Thus far in this dissertation, the emphasis has been on CW measurements. However, as will be demonstrated shortly, the reconstruction process for TOF measurements is almost identical to that for CW. The only difference is that the pixel values in the reconstructed image will be for index of refraction (square root of permittivity), instead of for attenuation.

Some of the reasons for considering both types of measuremnt systems include:

a) CW measurements give information on the attenuation of the earth between boreholes. The attenuation constant, whose formula was given in (2-5), is a function of the permittivity and the conductivity (we assume the permeability and the frequency to be constant). As will be shown, TOF measurements can give information on permittivity. Therefore, by combining both types of measurements, the permittivity and conductivity of the earth can be determined.

b) Because of diffraction effects, CW measurements can lead to ambiguous interpretation of reconstructed images. For example, as shown in Fig. 6.1, the shapes of the received

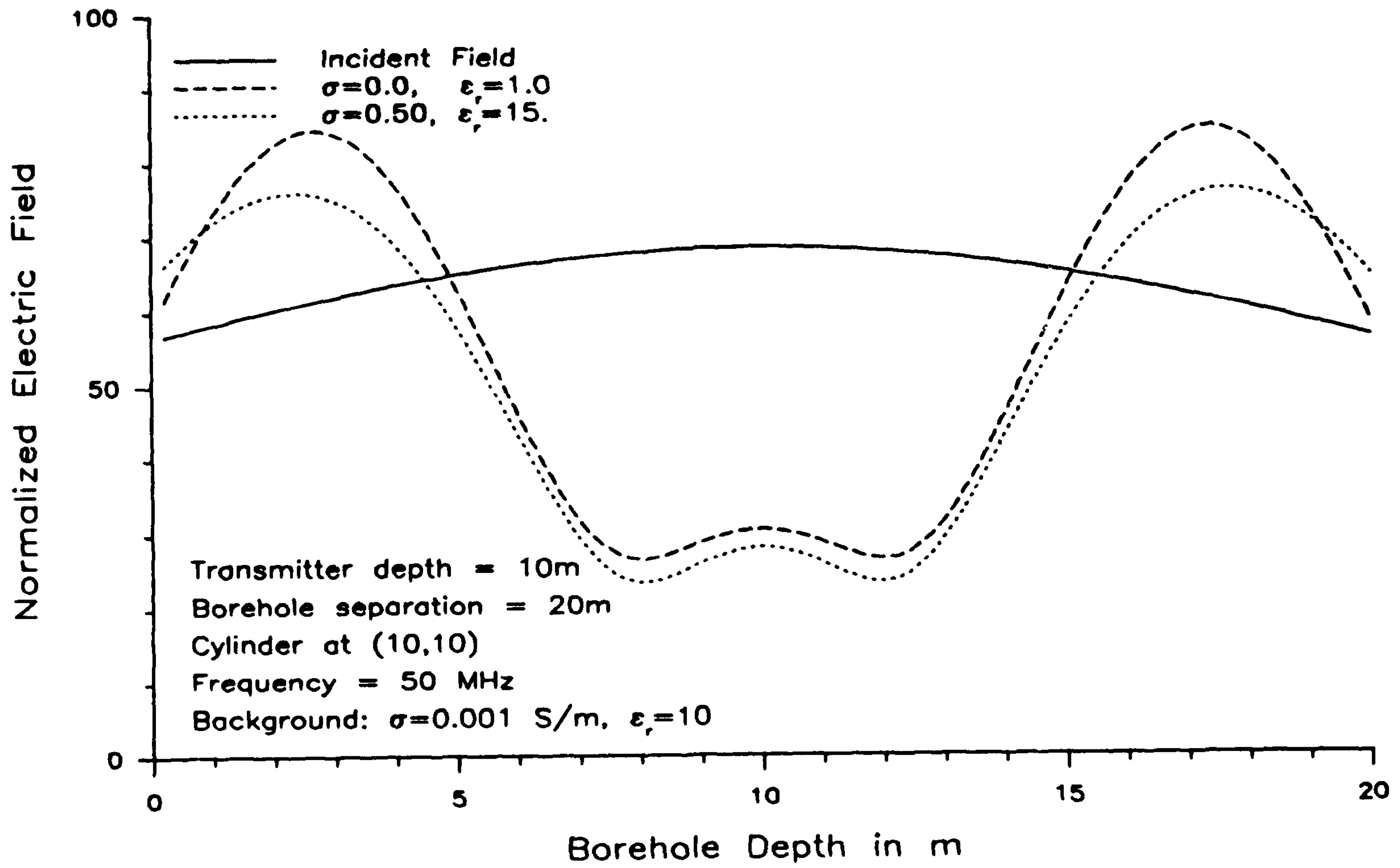


Fig. 6.1. Magnitude response of a tunnel and a conducting cylinder in a homogeneous earth.

electromagnetic field versus borehole depth when the earth contains a tunnel and when it contains a high conducting cylinder are very similar. Supplementing CW measurements with TOF measurements will help remove this ambiguity.

c) Some of the errors in each measurement system are unique to that system, and therefore by combining the two systems the errors may be reduced. For example, determining the time of arrival of a pulse in a TOF system involves the use of some kind of pulse picking process which will be prone to errors. The pulse picking process is an automated process for determining the time-of-arrival of the first pulse. A CW system will not have this problem.

For these and other reasons, it makes sense to obtain as much information about the region being scanned as is possible. Therefore, we first discuss reconstruction for TOF measurements, and then present a way of combining results from analyzing CW and TOF data.

6.2 Reconstructions from TOF Measurements

The electromagnetic wave velocity was given in (2-7). Using this wave velocity, the time it takes for a pulse to travel (assuming travel along straight ray paths) from a transmitting to a receiving antenna is given by the line integral

$$t = \int_L \frac{d\ell}{v(x,y)}, \qquad (6.1)$$

where L is the line linking the transmitter and the receiver. The wave velocity is a function of the permeability, permittivity, conductivity, and frequency which are assumed to be functions of positions (x and y). If the contribution of the conductivity is assumed to be negligible (refer to Figs. 2.7-2.9 for conditions under which this assumption is

valid) then the wave velocity is given approximately by

$$v = \frac{1}{(\mu\varepsilon)^{1/2}} = \frac{c}{\sqrt{\varepsilon_r}} \qquad (6\text{-}2)$$

where c is the speed of light in free space, and ε_r is the relative permittivity. Using this approximation, (6-1) becomes

$$t = \frac{1}{c}\int_L [\varepsilon_r(x,y)]^{1/2}\, d\ell. \qquad (6\text{-}3)$$

This equation can be discretized by again dividing the cross-sectional region into rectangular pixels and assuming the permittivity is constant over each pixel. A matrix equation is obtained as in (3-26), where the unknown image vector, x, is now composed of the square roots of relative permittivity in each pixel.

6.3 Detecting Anomalous Regions Using CW and TOF Measurements

Methods have been described for reconstructing a cross-sectional image of the earth. From this image we wish to detect, locate, and identify anomalous regions (if they exist) in the cross-section. The method we propose to use in this process can be summarized as follows:

a) Reconstruct the attenuation image from CW measurement data. Perform the image processing techniques described in Chapter V to paritition the image into low and high attenuation regions.

b) Perform the same steps as in step a
to the TOF data to obtain an image having low and high permittivity regions.

c) Compare the two partitioned images of steps a and b to determine if there is any overlap in the partitionings.

d) Using results on the electrical characteristics of earth (see,

for example, [24]) to determine the type of region that has been detected.

The exact procedure to be used in step d) is beyond the scope of this dissertation, but the interested reader may refer to [24] and the references therein for further information.

An example of using this four step process for detecting a tunnel will be given. It is assumed that a 1 m radius tunnel is located between two boreholes. The region being scanned is 20m x 20m and the background conductivity and relative permittivity are 0.001 S/m and 10, respectively. The increment between antenna locations (for both transmitter and receiver) is 1 m for a total of 400 measurements.

The MMP displays for both the CW and TOF reconstructions are shown in Fig. 6.2. In the (a) part of the figure, the higher attenuating regions are indicated by darker shading, while the darker shading in (b) indicates higher permittivity regions. The eight pixel region in the center of the image is seen to have higher attenuation yet lower permittivity. This clearly indicates the presence of a tunnel since a region of higher conductivity (e.g. a water deposit) would also have a higher permittivity. Thus, the ambiguity shown in Fig. 6.1 has been resolved.

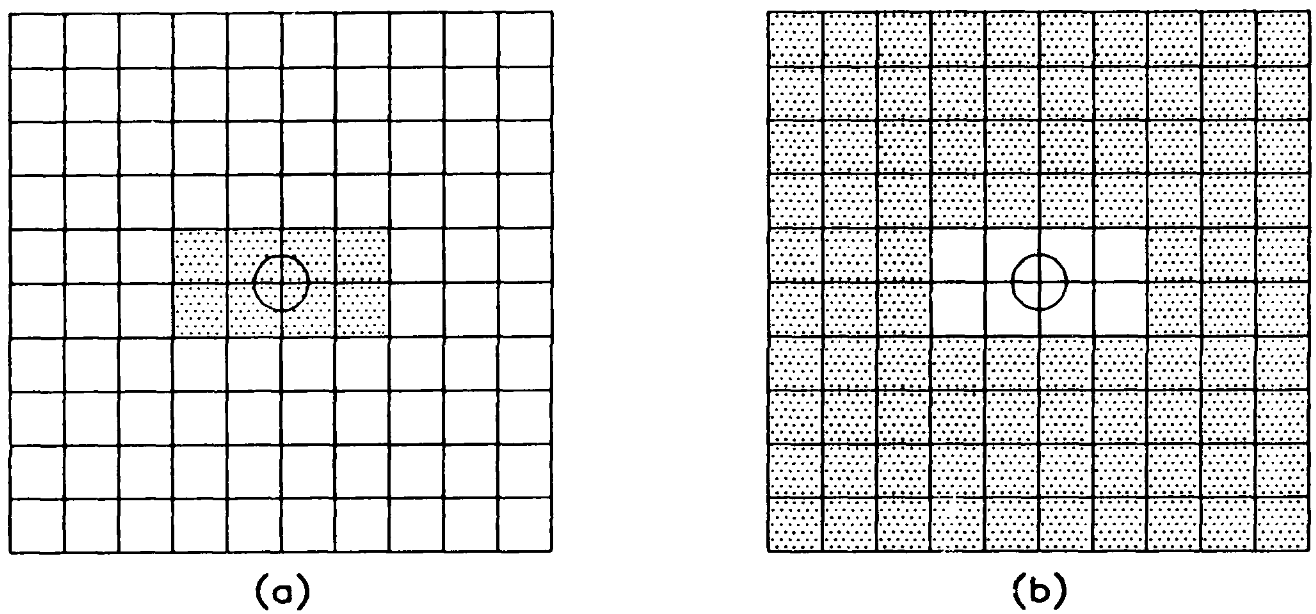


Fig. 6.2. Reconstructions for a tunnel in a homogeneous earth. (a) Reconstruction using CW measurements, darker shading represents higher attenuation. (b) Reconstruction using TOF measurements, darker shading represents higher permittivity.

CHAPTER VII

RECONSTRUCTIONS

7.1 Introduction

In this chapter the methods developed in the preceding chapters will be used to reconstruct cross sectional images from simulated data. We will be interested in cross sections of homogeneous earth containing single or multiple anomalies. The goal is to detect, identify, and locate these anomalies. One interest will be in comparing the algorithms developed in Chapter IV, under adverse conditions, such as limited view angles, noisy data, and multiple anomalies.

7.2 Simulation of Data

The simulation of data in a cross-hole environment was discussed in detail in Chapter II. For CW operations either exact solutions for circular cylindrical anomalies (discussed in Section 2.3.2) or the volume current method for arbitrarily shaped anomalies (discussed in Section 2.3.3) will be used to simulate the data. For TOF operation the ray optics method (discussed in Section 2.4) will be used to find the rays linking transmitters and receivers. From this ray path length information, the time it takes the ray to traverse the path is easily obtained from the electrical parameters of the earth along the path. For cases in which rays can reach the receiver along different paths, the path which results in the shortest travel time will be used. The method discussed in Section 2.3.5 (using the FFT) could be used for this simulation, but since we are only interested in the time it takes for the earliest pulse to reach the receiver, the ray optics method, discussed in Section 2.5.2 will be adequate (and computationally

faster).

7.3 Review: Total Geotomography Process

In this dissertation a variety of methods were discussed for reconstructing an image from its projections. In this section a review of the steps that will be taken to reconstruct a cross section of the earth from cross-hole data will be given. The emphasis will be on the detection of anomalous regions, thus the geotomography process will be geared to this problem.

Fig. 7.1 is a block diagram of the reconstruction process that will be used in this chapter. Note that there are 'inputs' for both CW and TOF data. The reconstruction algorithm featured in one of the blocks will consist of assuming a straight ray model as in Section 3.2.1 and then using one of the algorithms developed in Chapter IV. Methods for improving upon the straight ray model will be discussed in Section 7.6. The output of the reconstruction algorithm will be an image of attenuation or the index of refraction (square roots of permittivity) values.

This image is then processed using the techniques of Chapter V. This processing includes using 1 iteration of selective smoothing filtering and then partitioning the filtered image using the MMP. If anomalous regions are clearly identifiable in the processed image, then the images from both the CW and TOF data are compared as in Chapter VI. If at this point an anomalous region has been identified, we can attempt to refine the exact location (the images give only an approximate location due to the finite resolution of the pixels) of the anomaly in the region by using refinement methods which will be discussed in Section 7.6. Note that if an anomalous region has been found in either the attenuation or index of refraction image, but not in both, then both data sets are suspect, and further study would be required.

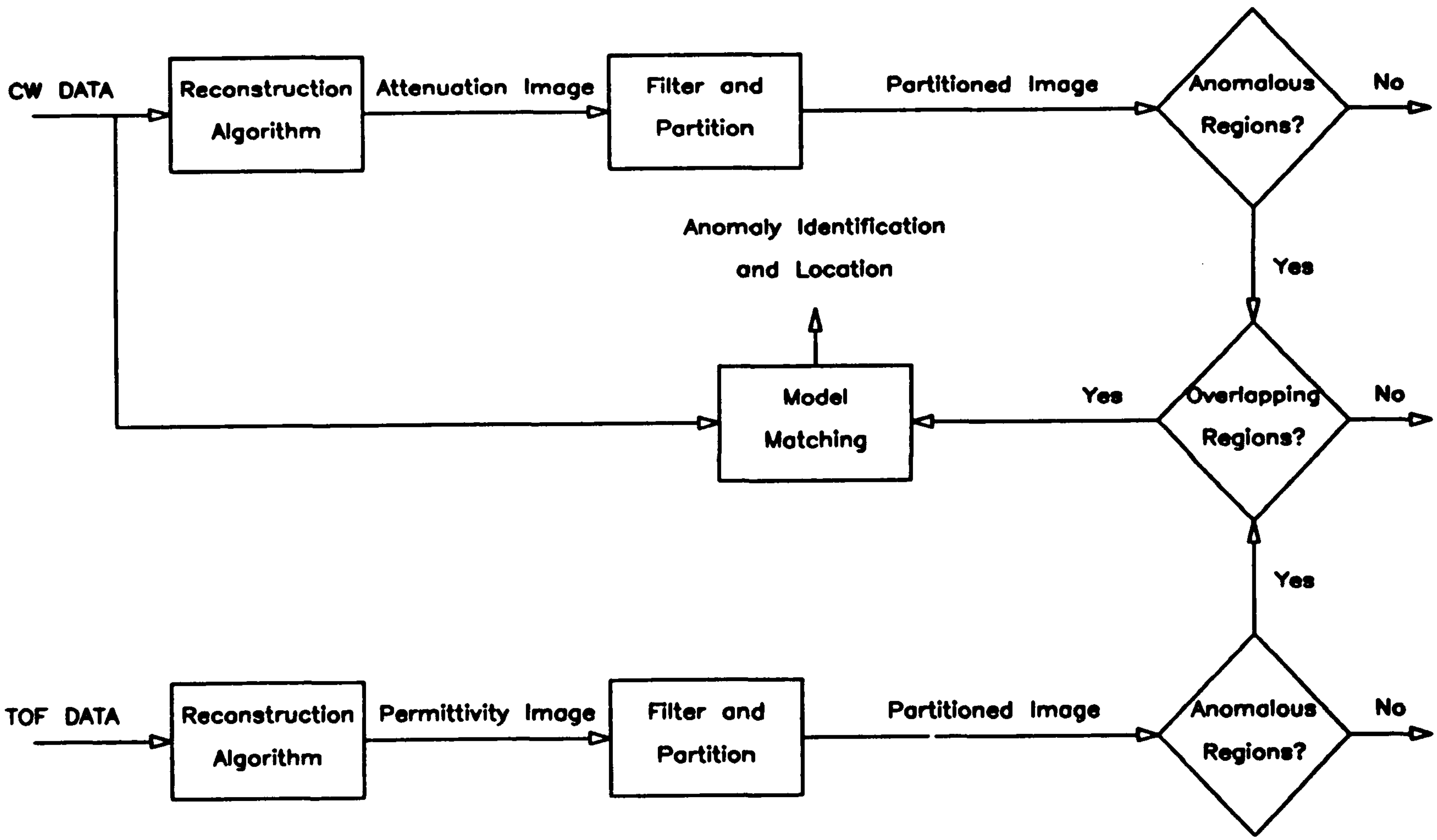


Fig. 7.1. Block diagram illustrating the cross-borehole reconstruction process.

The algorithms developed in Chapter IV were the singular value decomposition (SVD), the algebraic reconstruction technique (ART), and the gradient projection method using conjugate gradient minimization (CG-GPM). Unless otherwise noted, these algorithms will all be inverting the weighted normal equation given in (4-8), that is, we will be using the WLS method. A value of α=1.8 will be used for the path weighting in (4-10). All of the images will be constrained to have pixels whose values are greater than zero. The method for applying constraints was discussed in Chapter IV.

For the SVD algorithm of (4-16), γ will be chosen to damp out the effects of the small singular values as discussed in Chapter IV. The truncation value N will be selected in a similar manner (refer to Chapter IV for details). For the ART algorithm an underrelaxation value of 0.8 will be used in (4-28) as a compromise between faster converge and avoidance of cyclic behavior. This algorithm will also use Tikhonov regularization as in

$$(A^TWA + \gamma I)x = A^TWb, \tag{7-1}$$

with the regularization parameter, γ, chosen using the singular value decomposition as above. This equation is a result of combining (3-50) and (4-8).

The CG-GPM will also be used to solve (7-1) with the same regularization parameter found above. The stopping procedure discussed in Section 4.4.3 will be used to determine the number of iterations to be performed. However, this stopping point is not critical since the effects of the small singular values will be damped through use of the regularization parameter in (7-1). Iterating past this point will result in unnecessary computations.

7.4 Reconstructions Using Three Inversion Algorithms

7.4.1 Test Profiles

We will be performing reconstructions using the three algorithms discussed above on test profiles which present difficulties of various forms to the reconstruction process. The six profiles to be considered are shown in Fig. 7.2. The attributes of these profiles are summarized in Table 7.1. All profiles feature one or two cylindrical anomalies (of circular or square cross section) imbedded in a homogeneous earth. The background earth is assumed to have a conductivity of 0.001 S/m and a relative permittivity of 10. All profiles except for the last are 20 m in extent in both the horizontal and vertical directions. The last profile is 40 m in the horizontal direction and 20 m in the vertical direction. For both CW and TOF, measurements will be taken at 20 transmitter and 20 receiver locations, with a spacing of 0.5 m between locations. This will result in a total of 400 measurements. The size given in the second column of Table 7.1 is either the circle's radius or the side of the square. The range of view angles represent the minimum and maximum view angles obtained as the location of the transmitting antenna is changed.

Table 7.1

Test profiles for comparing the algorithms

Profile	Shape	Size	No. of anomalies	Matches pixels	View angle
(a)	circ.	1 m	1	No	43.5° - 50.8°
(b)	sq.	2 m	1	Yes	43.5° - 50.8°
(c)	sq.	2 m	1	No	43.5° - 50.8°
(d)	circ.	1 m	2	No	43.5° - 50.8°
(e)	circ.	0.5 m	1	No	43.5° - 50.8°
(f)	circ.	1 m	1	No	26.7° - 25.4°

The profiles listed in Table 7.1 have the following properties:

a) Profile (a) features a single large circular cylinder centered in the region. The circular cylinder should be easy to detect

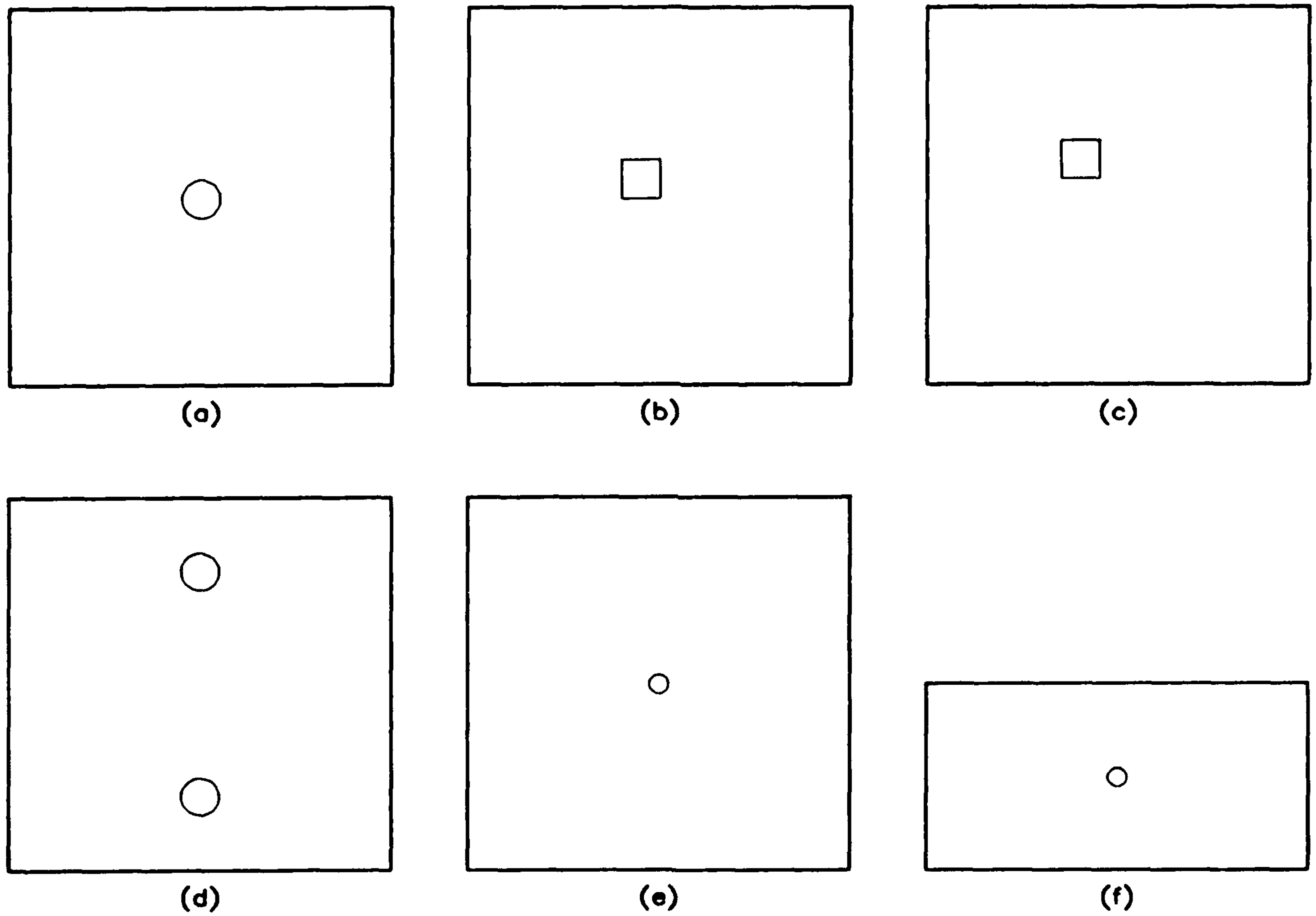


Fig. 7.2. Test profiles for comparing the reconstruction algorithms.

since it will be viewed over large angles.

b) Profile (b) features a single large square cylinder near the center of the region. The square cylinder is more difficult to detect since it will scatter the electromagnetic waves differently depending on the aspect angle. However, this cylinder coincides exactly with one of the square pixels used in the reconstruction process. This should allow this cylinder to be easily detected. In general, one should not expect the cylinder to coincide exactly with one of the reconstruction pixels.

c) Profile (c) is again a single large square cylinder near the center of the region. This time the cylinder does not exactly overlap a reconstruction pixel.

d) Profile (d) features two large circular cylinders centered between the boreholes. These cylinders will be difficult to detect since they do not lie near the center of the region where they would be subject to a maximum number of projections.

e) Profile (e) features a small circular cylinder centered in the region. This cylinder will be difficult to detect due to its small size (it is smaller than one wavelength at 50 MHz).

f) Profile (f) features a circular cylinder in a profile that is expanded in the horizontal direction. The limited view angle will make this cylinder difficult to detect.

These six profiles will provide a variety of conditions under which to test the various algorithms. It should be noted, however, that the algorithms will be able to reconstruct profiles which contain more complex anomalies than those being considered.

7.4.2 Reconstructions for High Conductivity Anomalies

In this section reconstructions will be obtained for the six

profiles described above for which the anomalous cylinder will have a higher conductivity (and permittivity) than the background medium. In all cases, the conductivity of the cylinder(s) will be 0.05 S/m with a relative permittivity of 20. The displayed results will be three-group (high, medium, and low attenuation) partitioned images of the reconstructed cross section for the CW simulations, and also three-group (high, medium and low permittivity) partitioned images for the TOF simulations. Ideally the (actual) location of the anomaly will coincide with the high attenuation (permittivity) region in the partitioned image.

The results for CW data will be presented first. For all the CW reconstructions, noise will be added to the simulated data to obtain a signal-to-noise ratio equal to 30 dB.

The reconstructions for the six profiles using the SVD algorithm are shown in Fig. 7.3. In the images, the actual locations of the anomalies are illustrated by the circles or squares drawn in the images. The darker shading indicates regions of higher attenuation. As can be seen from the figure, the SVD algorithm gives good reconstructions for almost all of the profiles. However, the small circular cylinder (profile e) is not located in the high attenuation region. In addition, extraneous pixels are included in the high attenuation regions in profiles (b), (d), and (f) of the figure.

The reconstructions in Fig. 7.4 were obtained by using 400 iterations of the ART algorithm. These images are similar in quality to the results shown in Fig. 7.3 in that in only one case (profile b) the cylinder does not lie in the high attenuation region, and some of the other cases (i.e., c, d, e, and f) have extraneous pixels in the high attenuation regions.

Fig. 7.5 shows reconstructions from using 180 iterations of the CG-GPM algorithm. For all cases the anomaly is located in the high attenuation region, and only three of the images (i.e., profiles b,d, and e) have extraneous pixels. This algorithm gives the best results

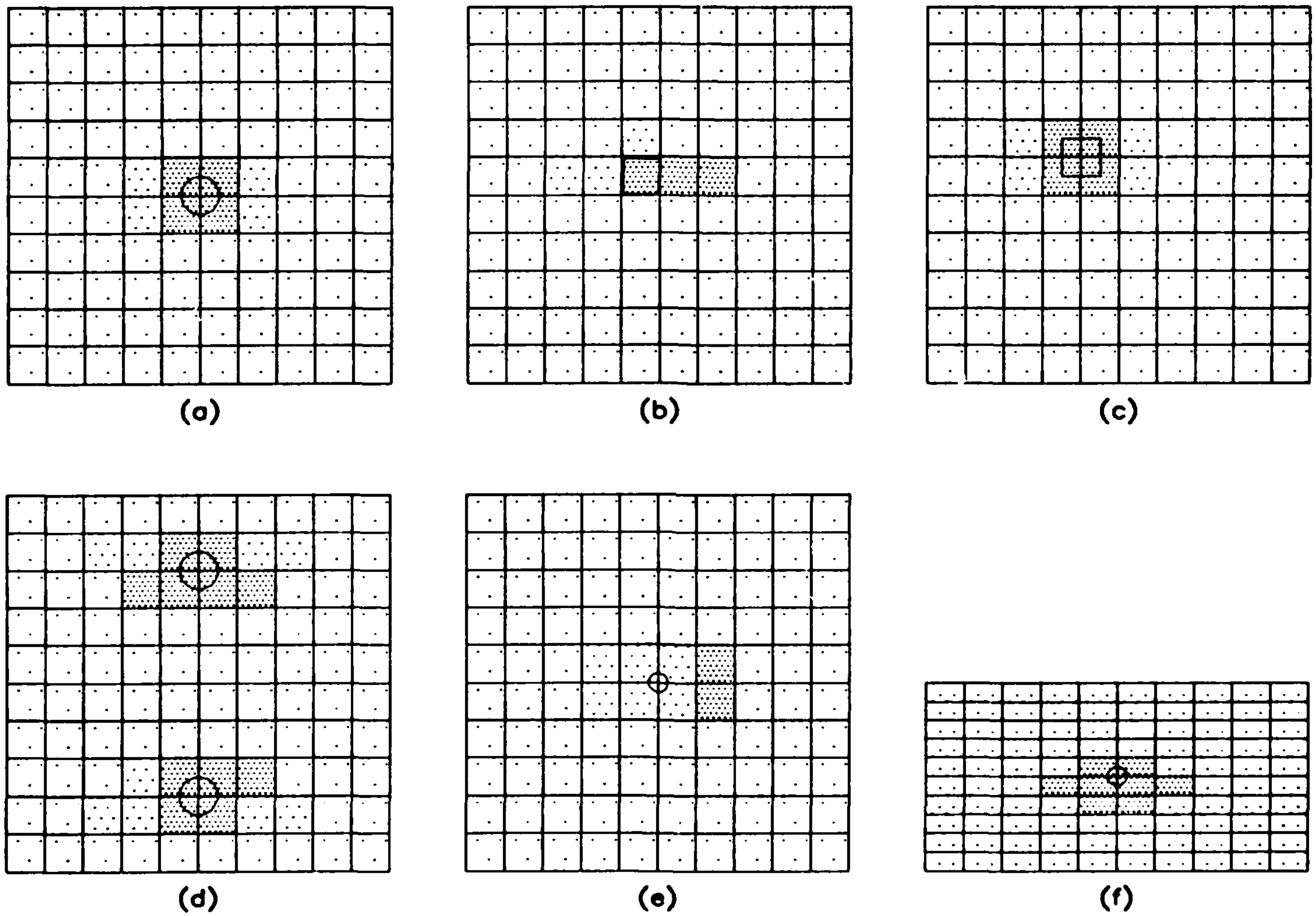


Fig. 7.3. Reconstructed images for the profiles listed in Table 7.1, high conductivity anomalies, CW data, SNR = 30 dB, SVD algorithm.

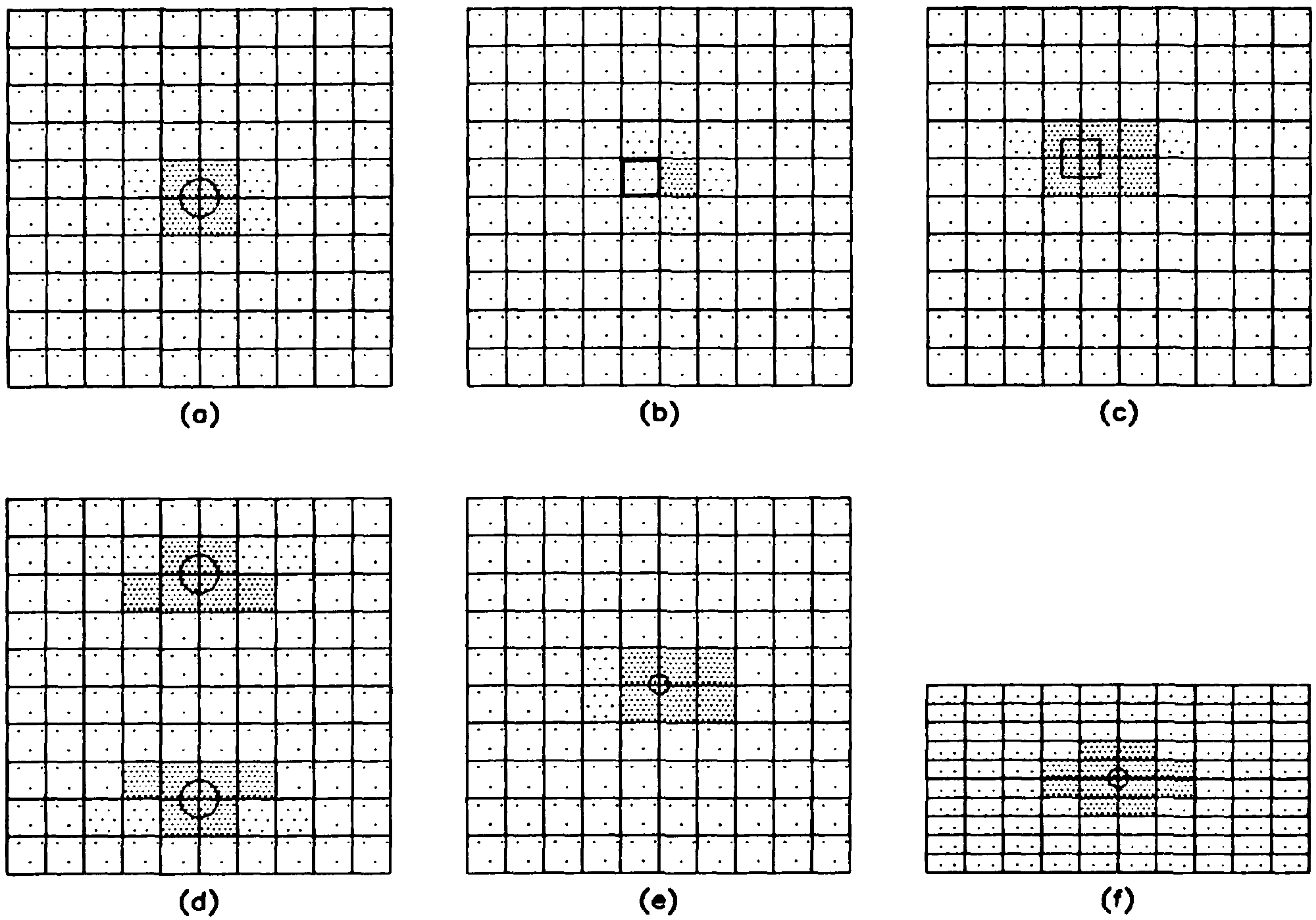


Fig. 7.4. Reconstructed images for the profiles listed in Table 7.1, high conductivity anomalies, CW data, SNR = 30 dB, ART algorithm.

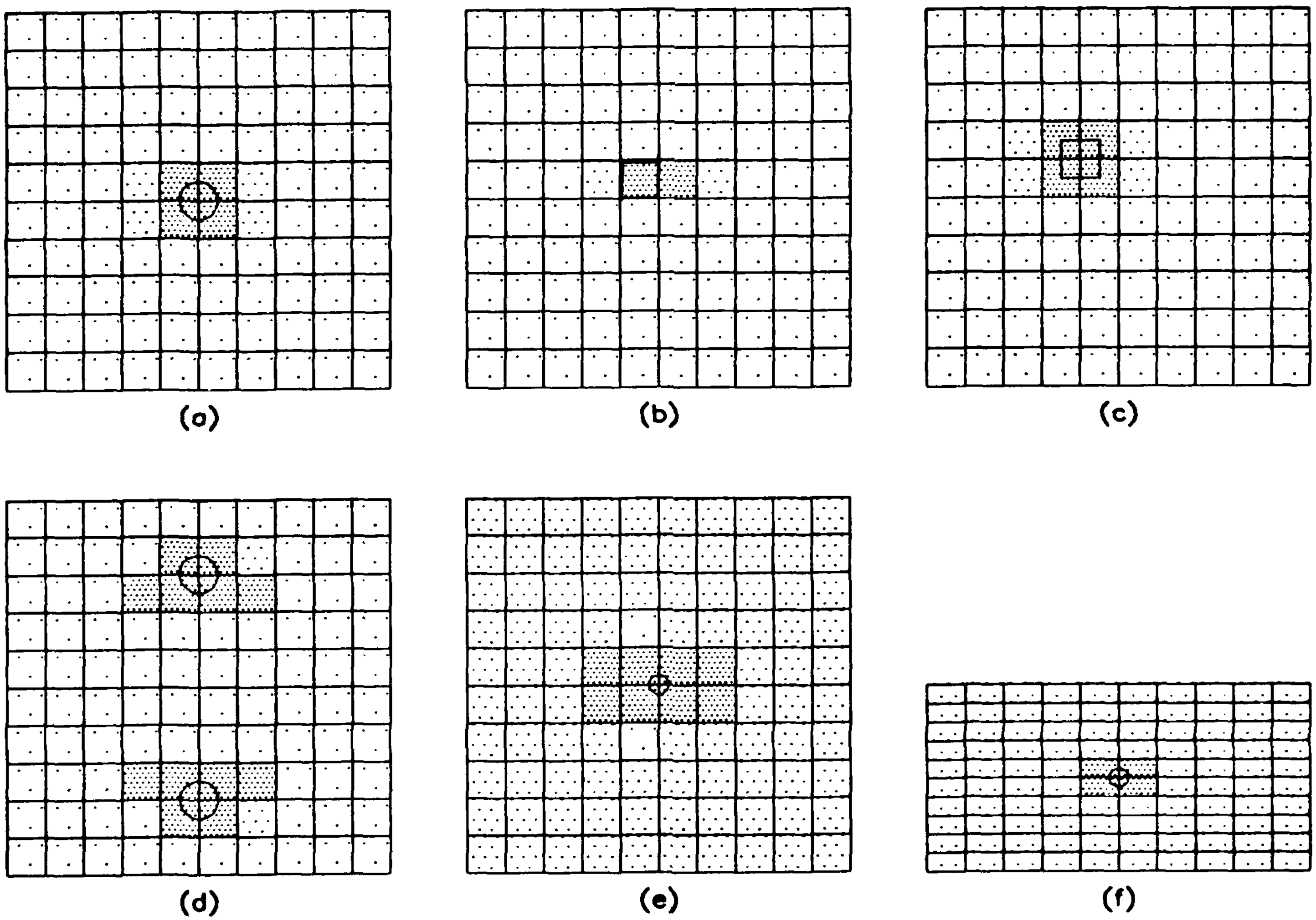


Fig. 7.5. Reconstructed images for the profiles listed in Table 7.1, high conductivity anomalies, CW data, SNR = 30 dB, CG-GPM algorithm.

for the CW data.

From these results, the following general conclusions for the CW reconstructions can be drawn.

a) In all of the reconstructions, the anomalies were detected. The only differences are in how accurately the anomalies were located and how many extraneous pixels are included in the high attenuation regions.

b) In all of the reconstructions there is spreading of the high attenuation region in the horizontal direction. This spreading is the result of the limited view angles for all of the reconstructions. There does not seem to be any way to avoid this phenomenon for the cross-borehole geometry.

c) In general, spreading in the vertical direction is limited to only a few of the reconstructions. This is due to the fact that there is adequate coverage in the vertical direction since the transmitter and receiver increments result in a fine spacing in this direction. This fine spacing is at the expense of requiring more measurements to be taken.

The results of using the SVD algorithm on the TOF data are shown in Fig. 7.6. For the TOF operation, 1% additive noise has been added to the simulated data. It is felt that in general TOF measurements will not be subject to as much noise since no calibration of the transmitted power will have to be made. As can be seen from the figure, profile (b) is the only one in which the cylinder is clearly identifiable. The reconstructions for profiles (a) and (d) show some indications of the presence of the cylinder(s), but the results are inconclusive.

ART reconstructions for the TOF data are shown in Fig. 7.7. In profiles (a) and (b) the anomalies are located in the high permittivity regions. However, extraneous pixels are present in profile (a). Profiles (c) and (f) have the anomalies partially in the high permittivity regions. Again, profile (d) shows some indications

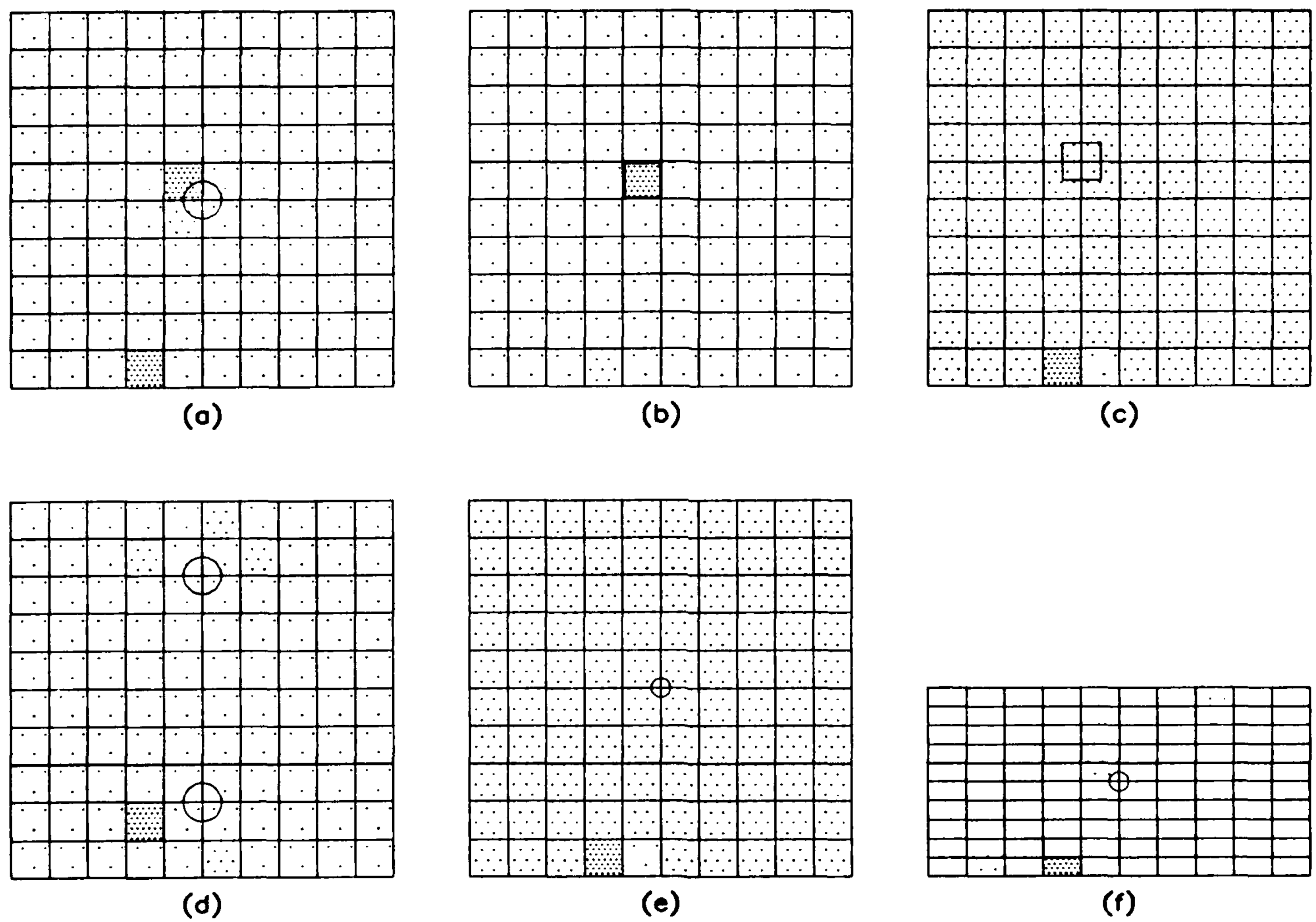


Fig. 7.6. Reconstructed images for the profiles listed in Table 7.1, high conductivity anomalies, TOF data, 1% noise, SVD algorithm.

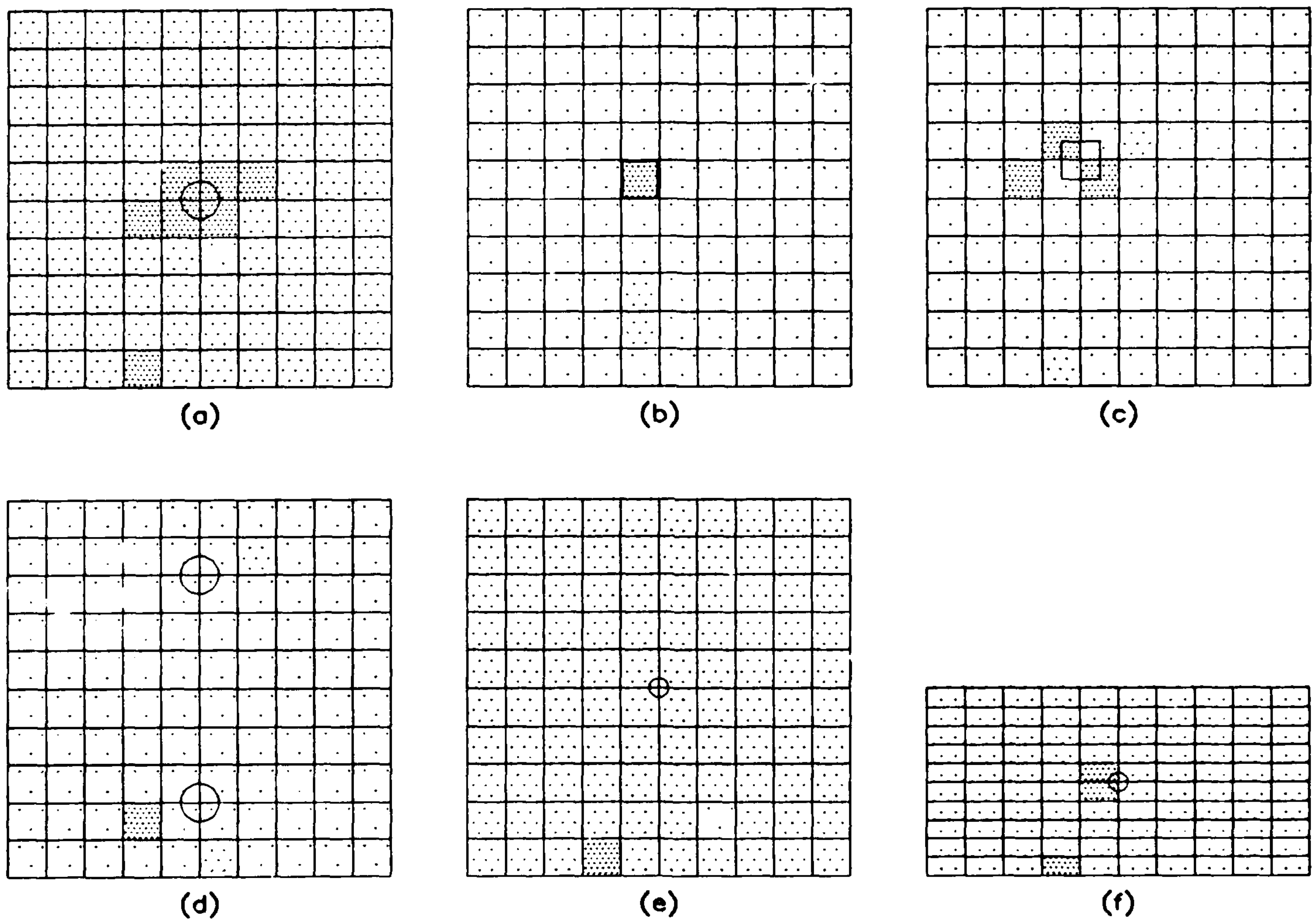


Fig. 7.7. Reconstructed images for the profiles listed in Table 7.1, high conductivity anomalies, TOF data, 1% noise, ART algorithm.

of the cylinders, but they are not easily identifiable.

Reconstructions using the CG-GPM algorithm are shown in Fig. 7.8. Again, the cylinders are located in the high permittivity regions for profiles (a) and (b). Profiles (c) and (f) have half of the cylinders in the high permittivity region, and some indication of the presence of the cylinders is evident in profile (d). Finally, although the cylinder is located in the high permittivity region for profile (e), the results are inconclusive. We can conclude from these results that for the TOF data, the CG-GPM algorithm gives the best reconstruction results. In addition, the following general conclusions for the TOF reconstructions can be drawn.

a) Unlike for the CW reconstructions, the anomalies were not detected in all cases. This gives further reason for obtaining both types of measurements.

b) When the anomalies were detected, the high permittivity region did not always contain the entire anomaly. This suggests that a partitioning scheme which would be biased to include neighboring pixels into the high permittivity region might perform better for TOF reconstructions.

It is worthwhile to investigate why the reconstruction results for the CW data are better than for the TOF data. To this end, Fig. 7.9 shows plots of the received electromagnetic field and pulse arrival times versus (receiver) borehole depth for profile (e). For this figure the transmitter is located at a depth equal to the center of the anomaly. Note that because of diffraction effects the cylinder 'shadow' for the received field magnitude (CW data) is larger than the optical (straight ray) 'shadow' of the cylinder. However, the attenuation of the field in this region is significant. On the other hand, for the time-of-flight data the 'shadow' from the cylinder is nearly the same as the optical shadow, but the difference between arrival times in the 'shadow' and 'lit' regions is very slight. This figure further emphasizes the need to obtain both CW and TOF

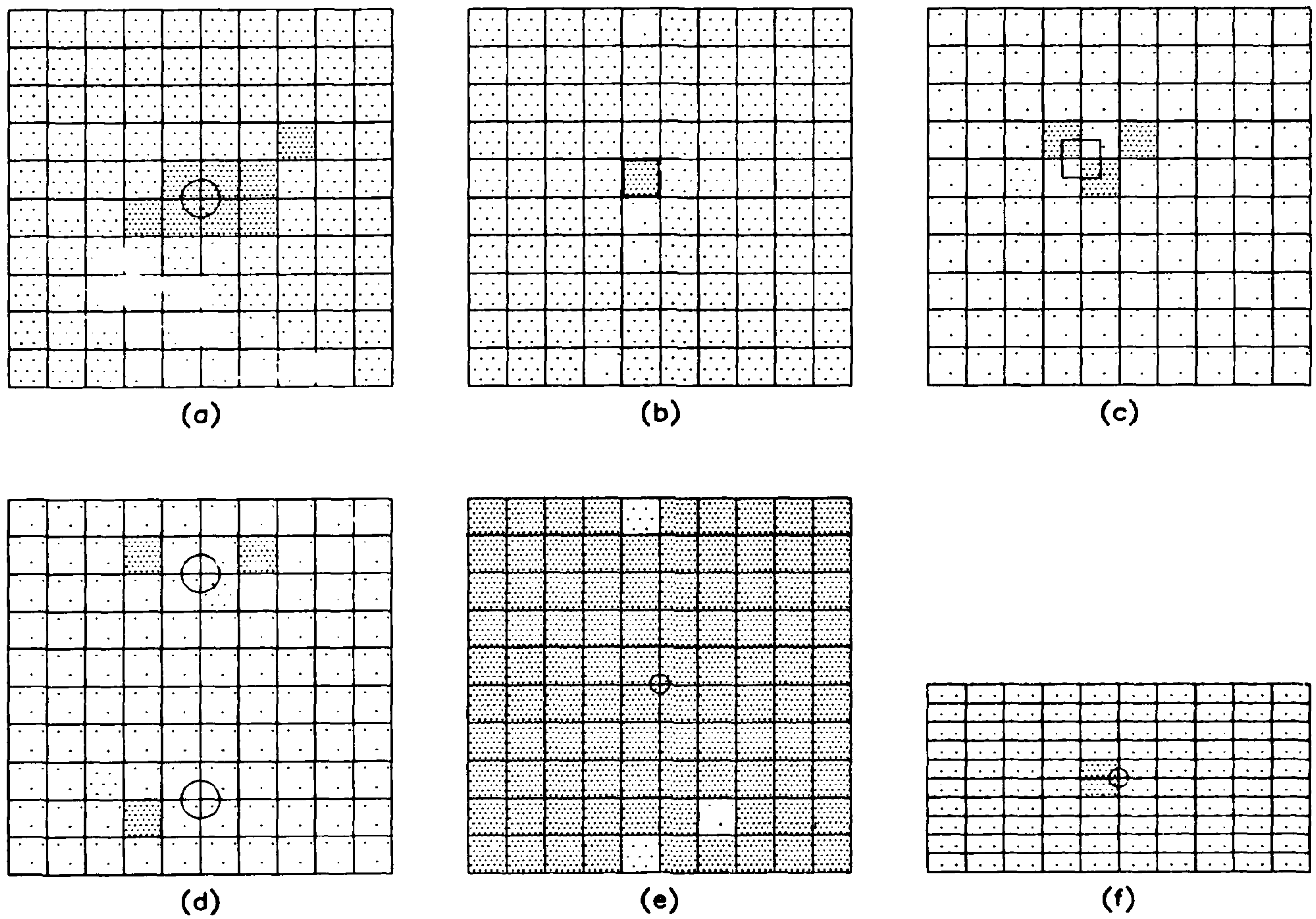


Fig. 7.8. Reconstructed images for the profiles listed in Table 7.1, high conductivity anomalies, TOF data, 1% noise, CG-GPM algorithm.

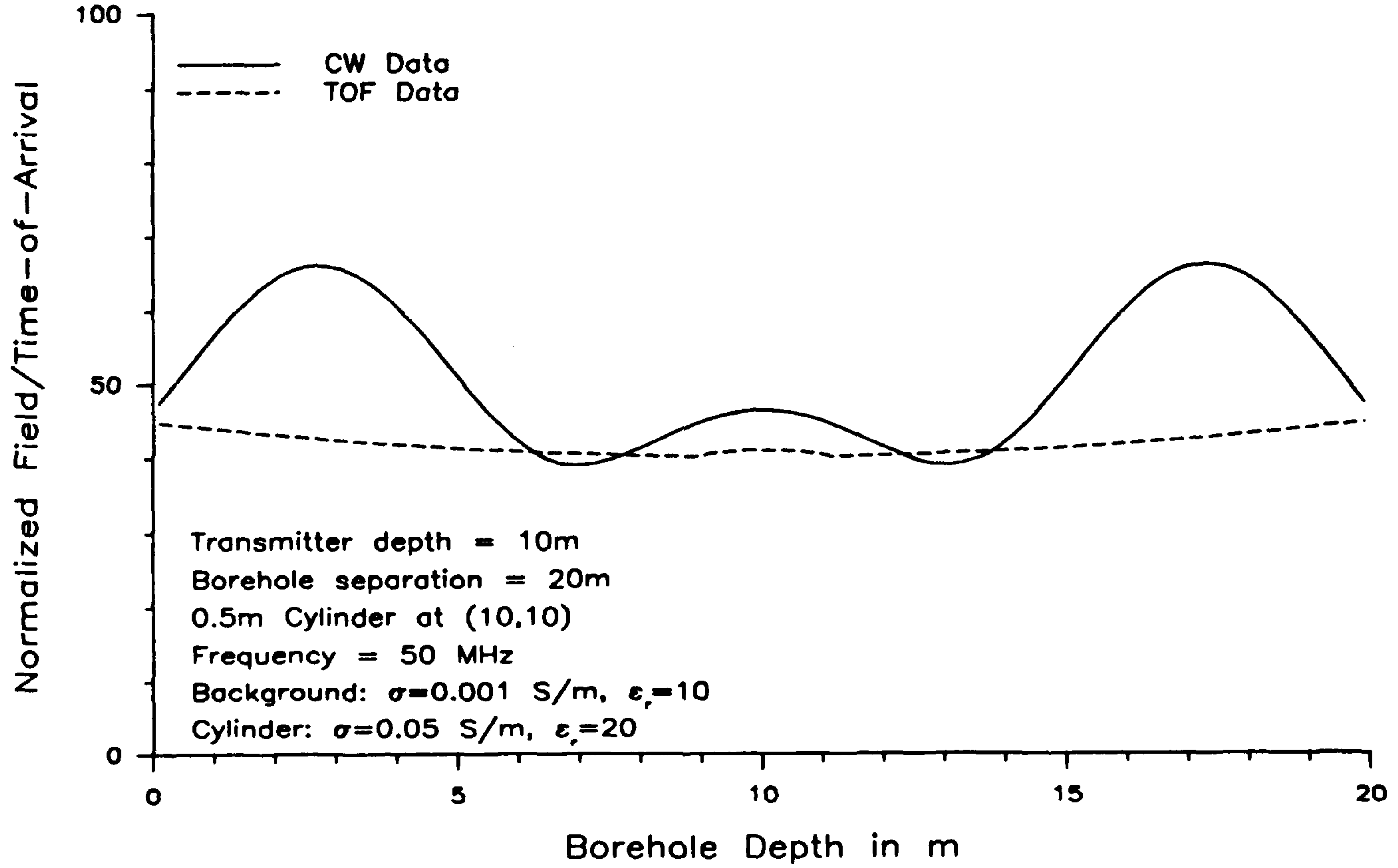


Fig. 7.9. Magnitude and time-of-flight response for 0.5 m tunnel in a homogeneous earth.

measurement data in that one set of data might yield better reconstructions depending on the profile.

The results of combining both CW and TOF reconstructions for the CG-GPM algorithm are shown in Fig. 7.10. This figure was obtained by intersecting the regions of high attenuation in Fig. 7.5 with the regions of high permittivity in Fig. 7.8 as suggested in Chapter VI. These results are shown for the CG-GPM algorithm since this algorithm gave the overall best reconstructions. All of the profiles except for (d) give a good indication of the location of the anomaly, although the reconstruction for profile (e) might be suspect since the TOF reconstruction (Fig. 7.8) for this profile did not give a localized region of high permittivity. For profiles (a), (b), (c), and (f) we can conclude that we have located regions of high attenuation and permittivity, which would signify, for example, a section of earth having high water content.

7.4.3 Reconstructions for Tunnels Located in the Earth

In this section the cylinder imbedded in the earth will be an air-filled void, that is, a tunnel. In this case the conductivity of the cylinder is zero and its relative permittivity is unity. As before, the conductivity of the background is 0.001 S/m and its relative permittivity is 10.

Reconstructions for CW data for the SVD, ART, and CG-GPM algorithms are shown in Figs. 7.11, 7.12, and 7.13, respectively. The results are similar to those obtained for the high conducting case. This is not surprising in light of the discussion in Chapter VI and the plots in Fig. 6.1. Because of the diffraction of the rays around the cylinder, the tunnel causes an attenuating effect (see Fig. 6.1), which results in a region of high attenuation in the reconstructed image. For the CW data it might be judged that the ART algorithm does slightly better than the CG-GPM algorithm from comparing the reconstructions for

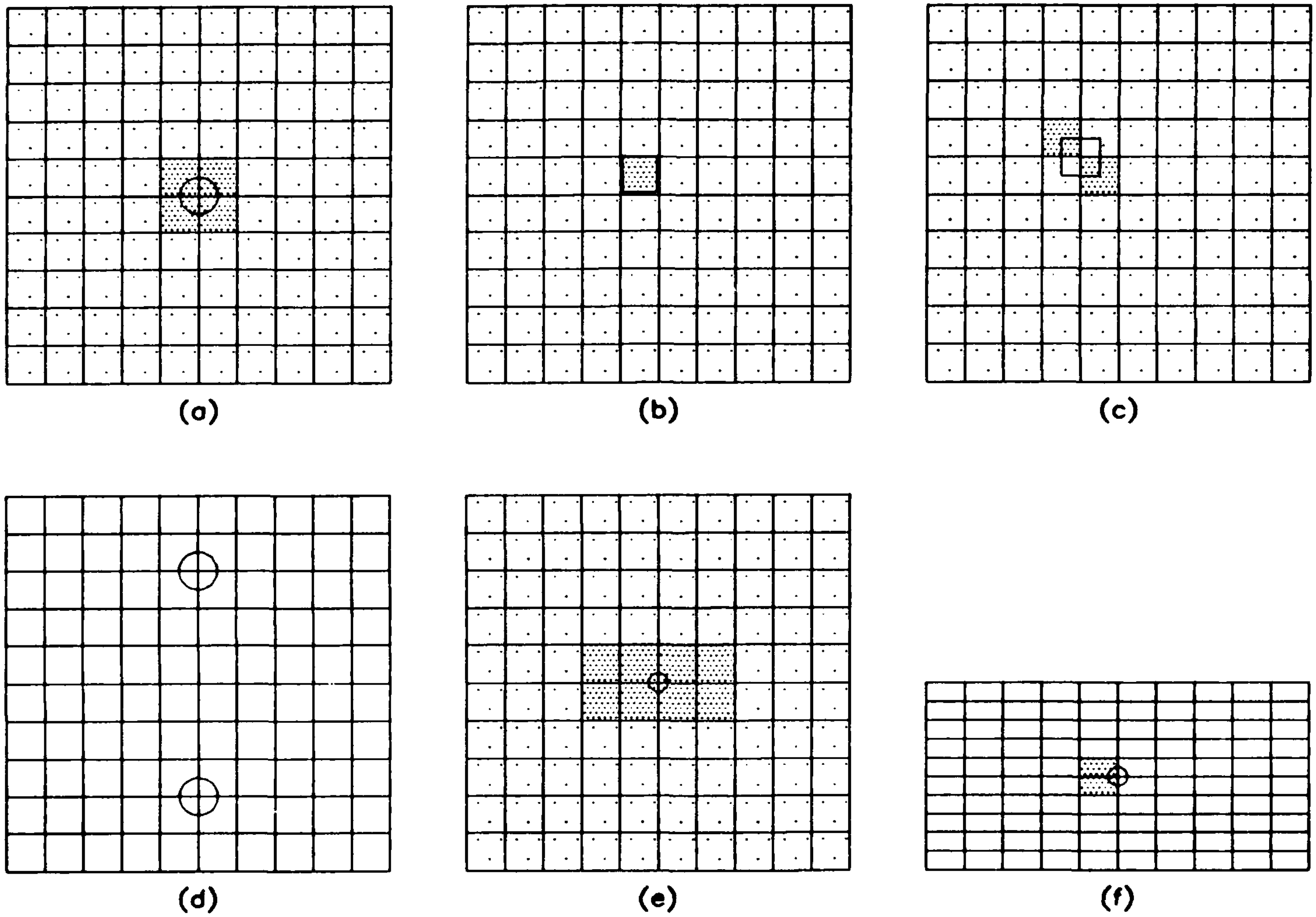


Fig. 7.10. Result of combining the CW and TOF reconstructions in Figs. 7.4 and 7.7 for the CG-GPM algorithm.

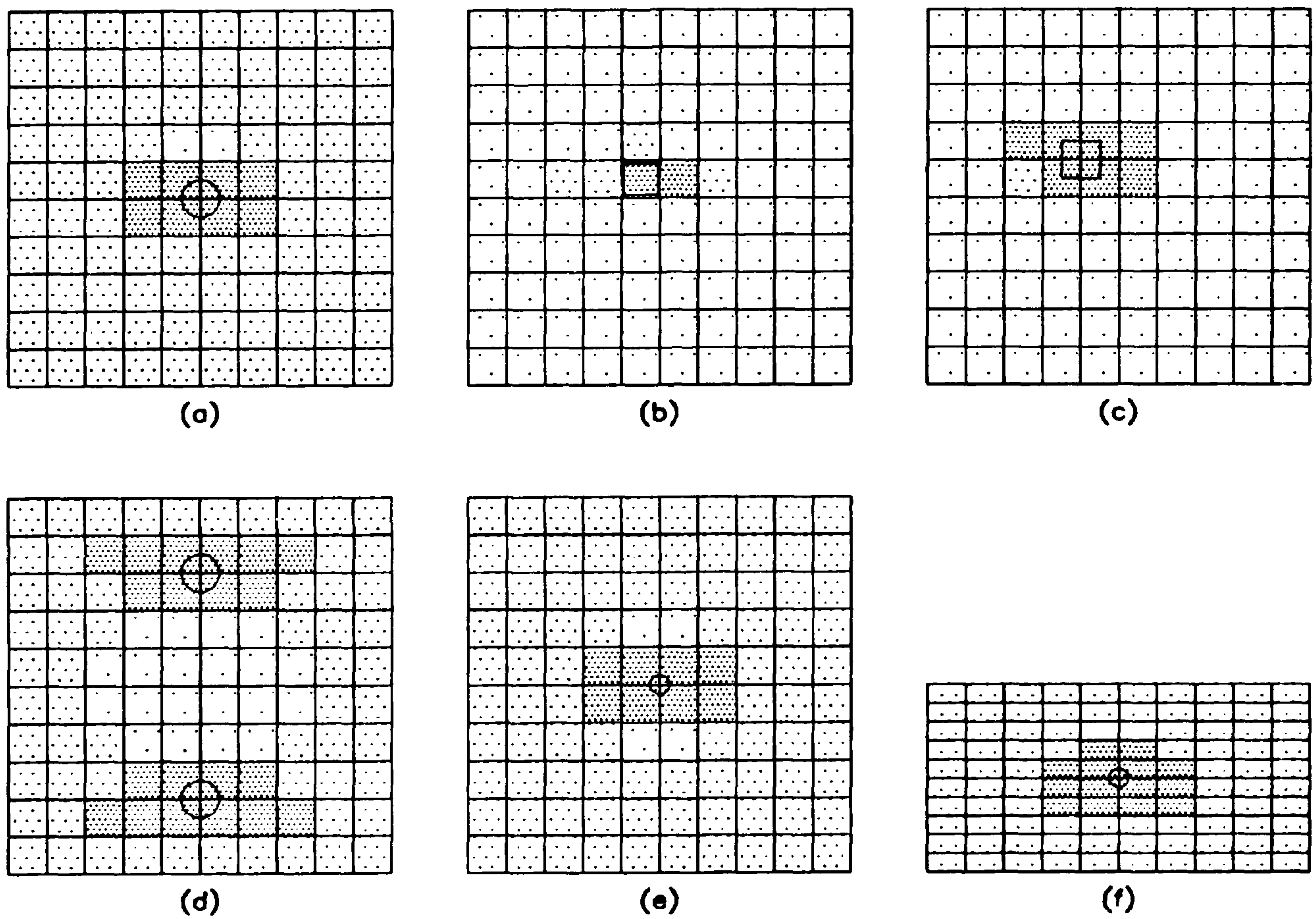


Fig. 7.11. Reconstructed images for the profiles listed in Table 7.1, low conductivity anomalies, CW data, SNR = 30 dB, SVD algorithm.

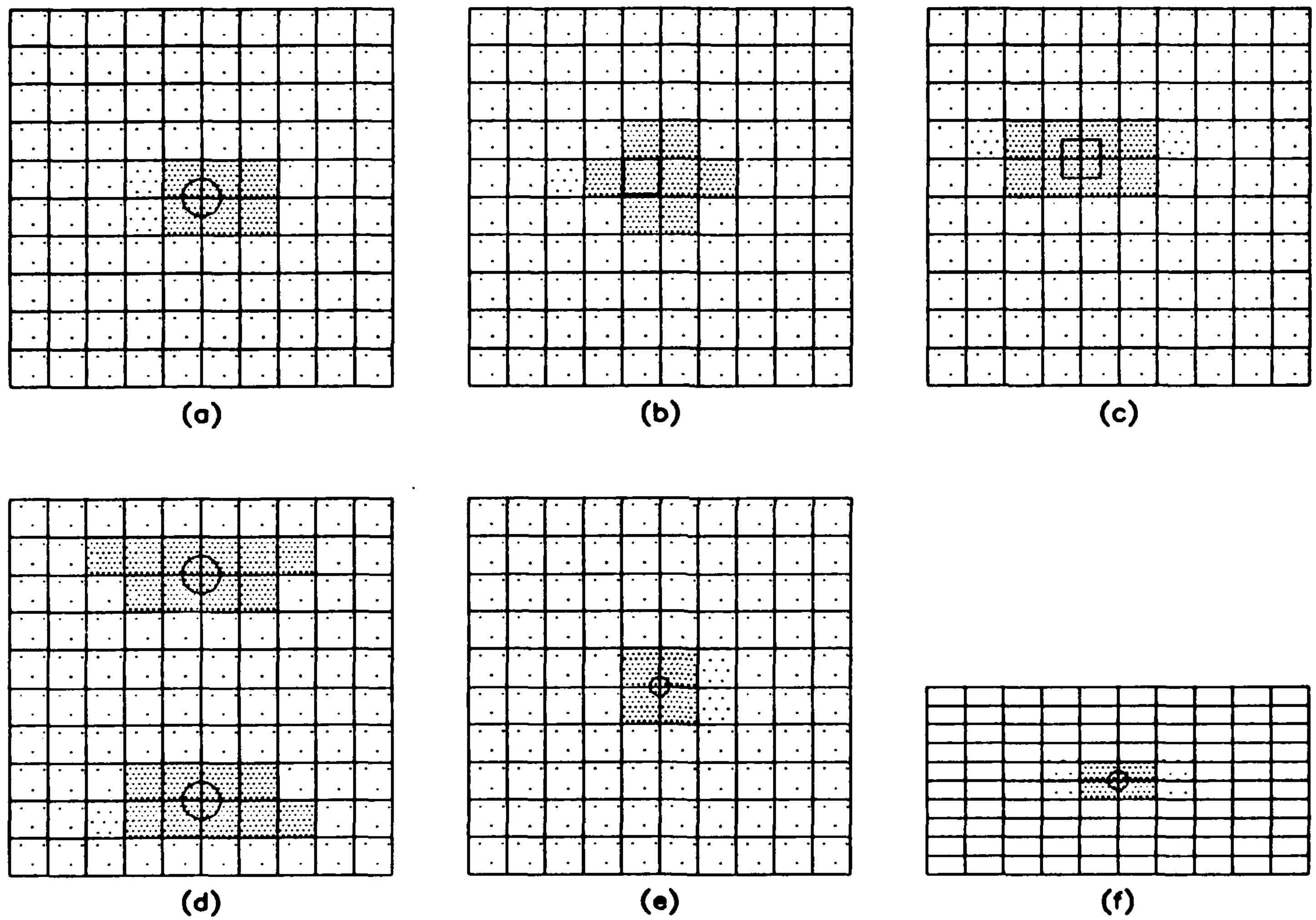


Fig. 7.12. Reconstructed images for the profiles listed in Table 7.1, low conductivity anomalies, CW data, SNR = 30 dB, ART algorithm.

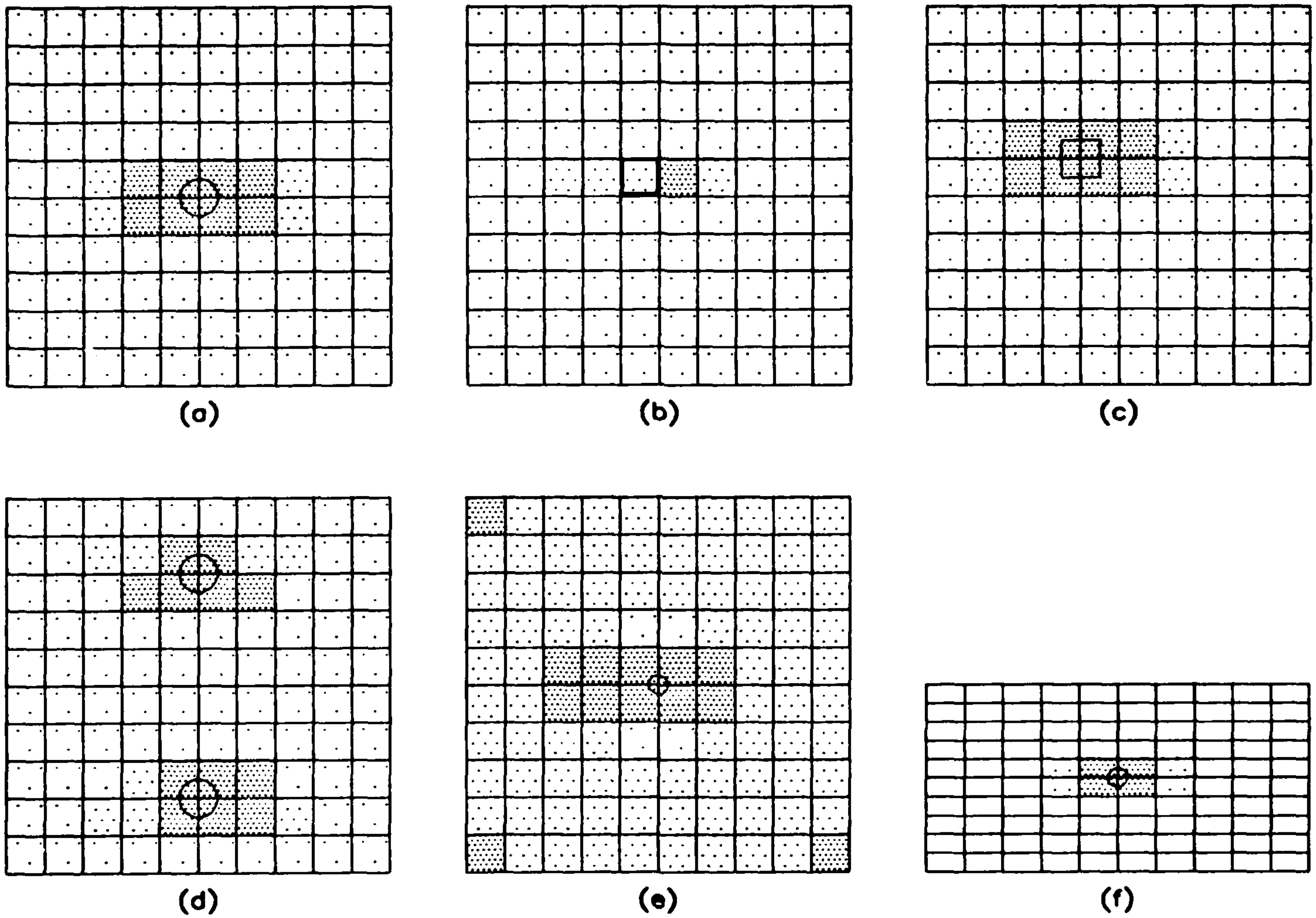


Fig. 7.13. Reconstructed images for the profiles listed in Table 7.1, low conductivity anomalies, CW data, SNR = 30 dB, CG-GPM algorithm.

profiles (b) and (e) in Figs. 7.12 and 7.13. However, for all the other profiles the performances of the two algorithms are very similar. The SVD algorithm gives reconstructions comparable in quality to the ART or CG-GPM, although there are many extraneous pixels in profile (f) of Fig. 7.11. In addition, the following general conclusions for the CW reconstructions can be drawn.

a) In all of the reconstructions, the tunnels were detected. Again, the only differences are in how accurately they are located and how many extraneous pixels are included in the high attenuation regions.

b) The spreading in the horizontal direction is again evident in all of the profiles.

c) The tunnel reconstructions are very similar to those obtained for the high conductivity anomalies. This was expected from the similarity of the magnitude responses of the tunnel and high conductivity cylinder in Fig. 6.1.

Reconstructions for TOF data are shown in Figs. 7.14, 7.15, and 7.16. Note that since the tunnels have lower permittivity than the background, the tunnels are identified by the regions that have light shading. All three algorithms are able to identify the tunnels in profiles (a) and (b), and all show parts of the tunnel in profiles (c) and (e). None of the algorithms is able to distinguish the two tunnels in profile (d), although the tunnels are located in the low permittivity region. Finally, only the CG-GPM algorithm is able to detect part of the tunnel in profile (f). In addition, the following general conclusions for the TOF reconstructions can be drawn.

a) Again, for the TOF reconstructions, the tunnels were not detected in all cases.

b) When the tunnels were detected, the low permittivity region was smaller (fewer extraneous pixels) than for the CW reconstructions. This phenomenon can be explained by referring again to Fig. 7.8.

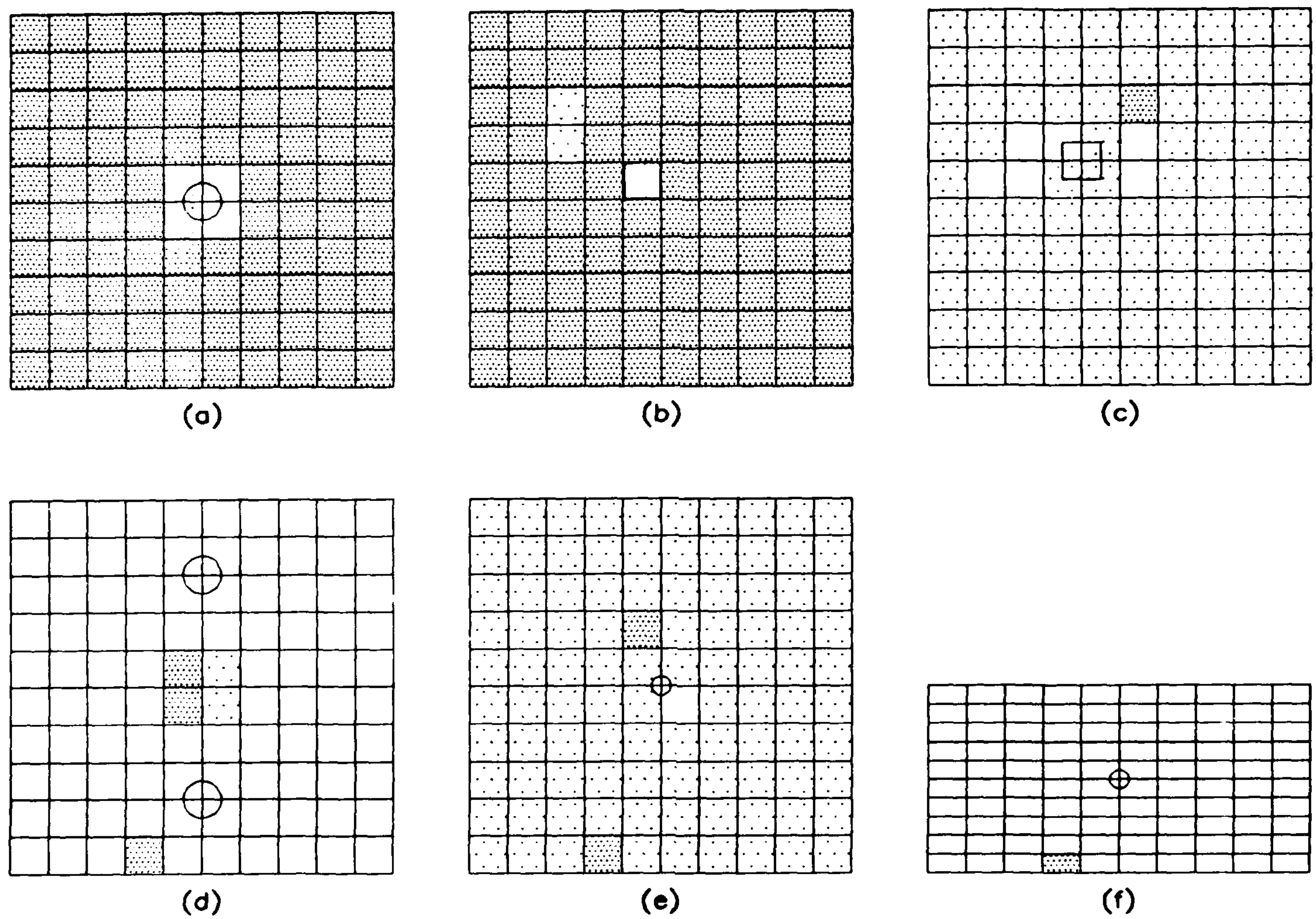


Fig. 7.14. Reconstructed images for the profiles listed in Table 7.1, low conductivity anomalies, TOF data, 1% noise, SVD algorithm.

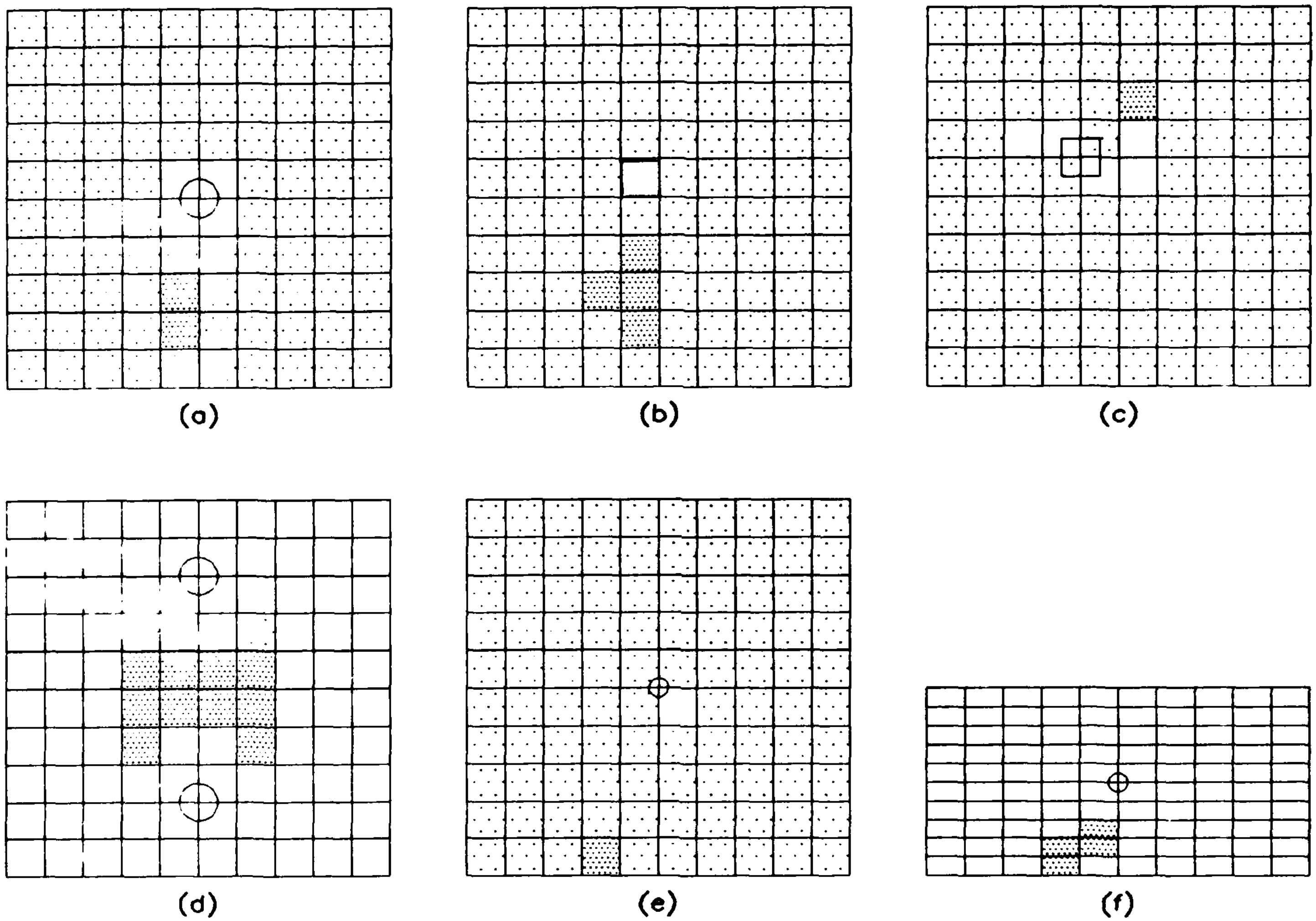


Fig. 7.15. Reconstructed images for the profiles listed in Table 7.1, low conductivity anomalies, TOF data, 1% noise, ART algorithm.

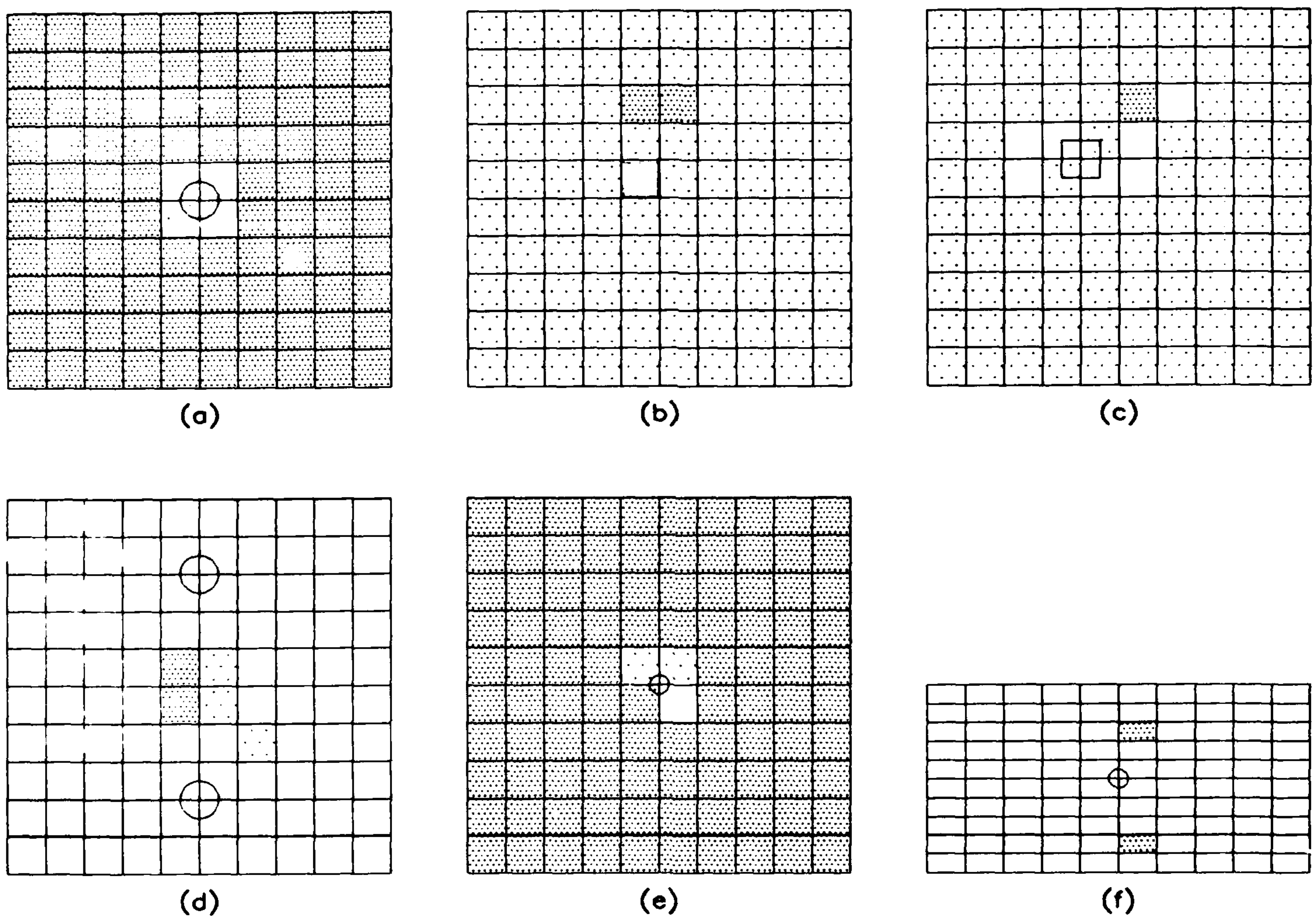


Fig. 7.16. Reconstructed images for the profiles listed in Table 7.1, low conductivity anomalies, TOF data, 1% noise, CG-GPM algorithm.

c) The reconstructions for the tunnels were in general superior to those obtained for the high conductivity anomalies. This is due to the fact that the ratio of the tunnel's permittivity to the background is 1:10, while for the high conductivity anomaly it was 2:1. This suggests that TOF measurements will in general be most effective for detecting tunnels.

Since it is unclear which algorithm has performed best in locating tunnels in the six profiles, all three have been used to combine CW and TOF data as described above. The results are shown in Figs. 7.17, 7.18, and 7.19. In these figures the dark shading indicates regions of high attenuation and low permittivity. From the discussion in Chapter 6, such regions would signify the presence of a tunnel. All three algorithms are able to identify the tunnels in profiles (a), (c), (d), and (e), although the results shown for the two tunnel case (profile d) are suspect since the TOF reconstructions for the profile were inconclusive. Both the SVD and ART algorithms located the tunnel in profile (b), but the CG-GPM was the only algorithm to identify the tunnel in profile (f). The CG-GPM did not identify the anomaly in (b) because the regions of high attenuation and low permittivity did not intersect. However, from studying the attenuation image of Fig. 7.13 (b) and the permittivity image of 7.16 (b) one would be led to believe that an anomaly exists in the region surrounding the actual anomaly location.

7.4.4 Conclusions

From the results presented in this section a number of conclusions may be drawn:

a) For the profiles considered, the CG-GPM performed better at locating high conducting anomalies than either the SVD or ART algorithms. This agrees with the results of Chapter IV where the CG-GPM algorithm performed best on all of the test cases.

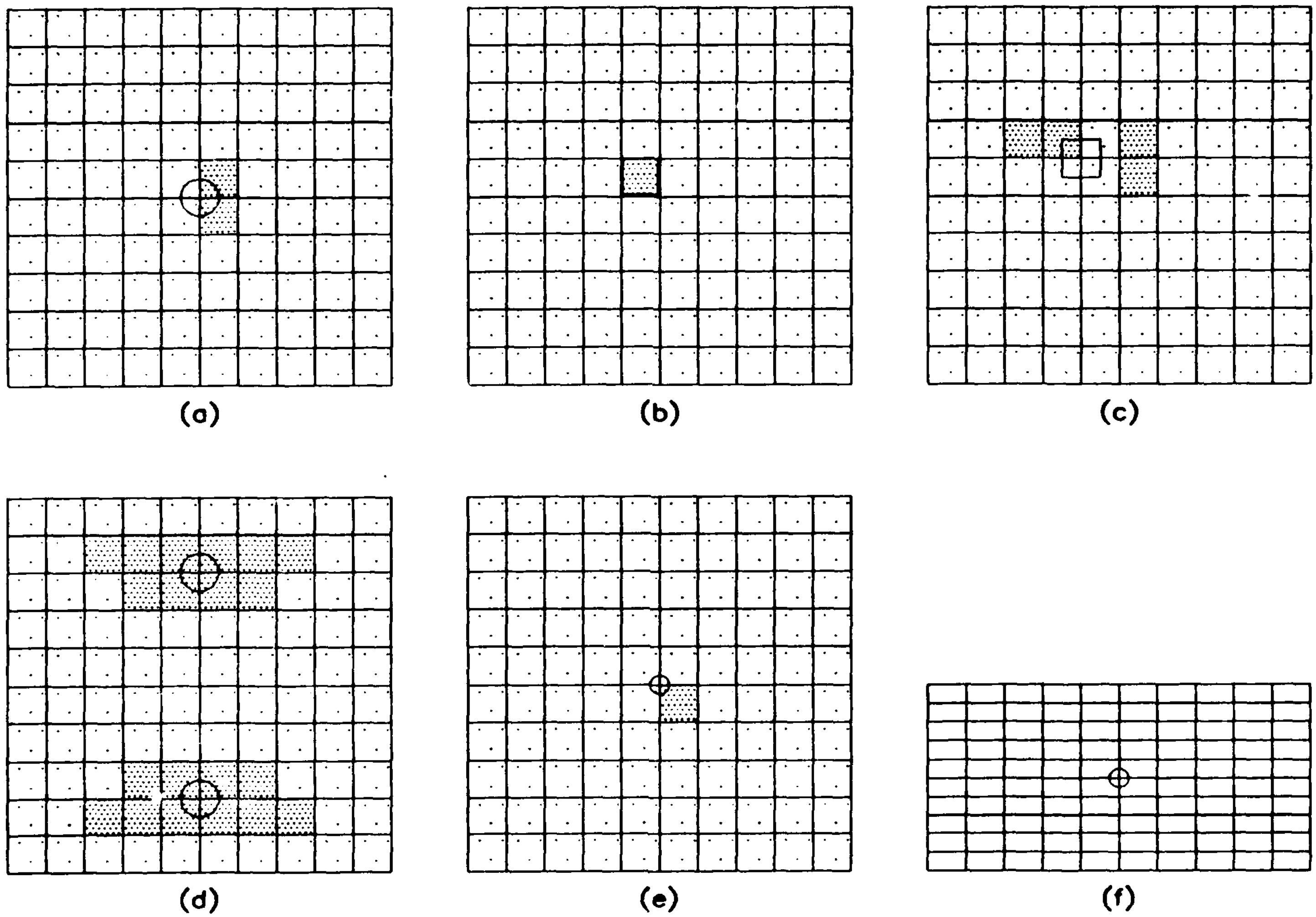


Fig. 7.17. Result of combining the CW and TOF reconstructions in Figs. 7.10 and 7.13 for the SVD algorithm.

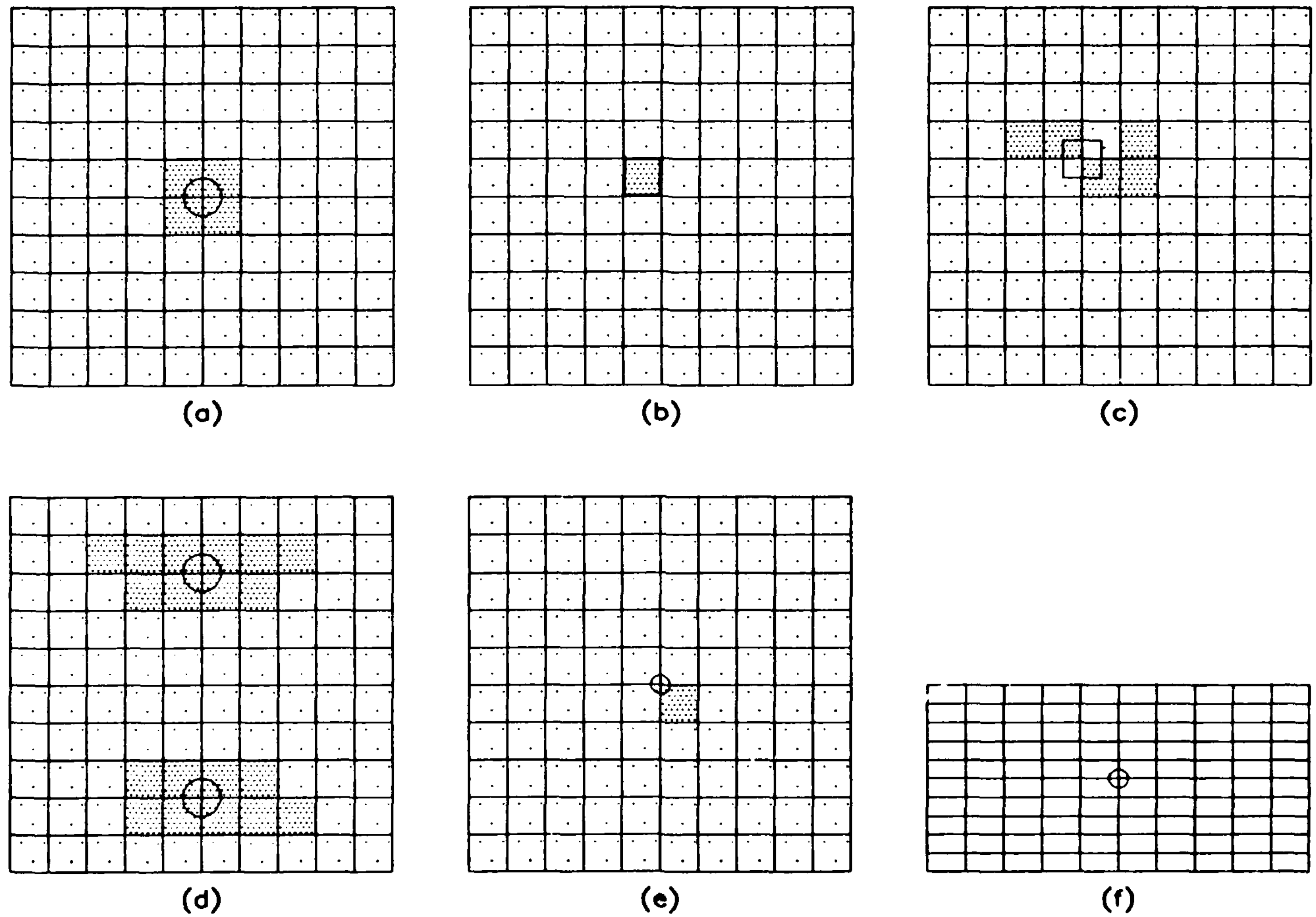


Fig. 7.18. Result of combining the CW and TOF reconstructions in Figs. 7.11 and 7.14 for the ART algorithm.

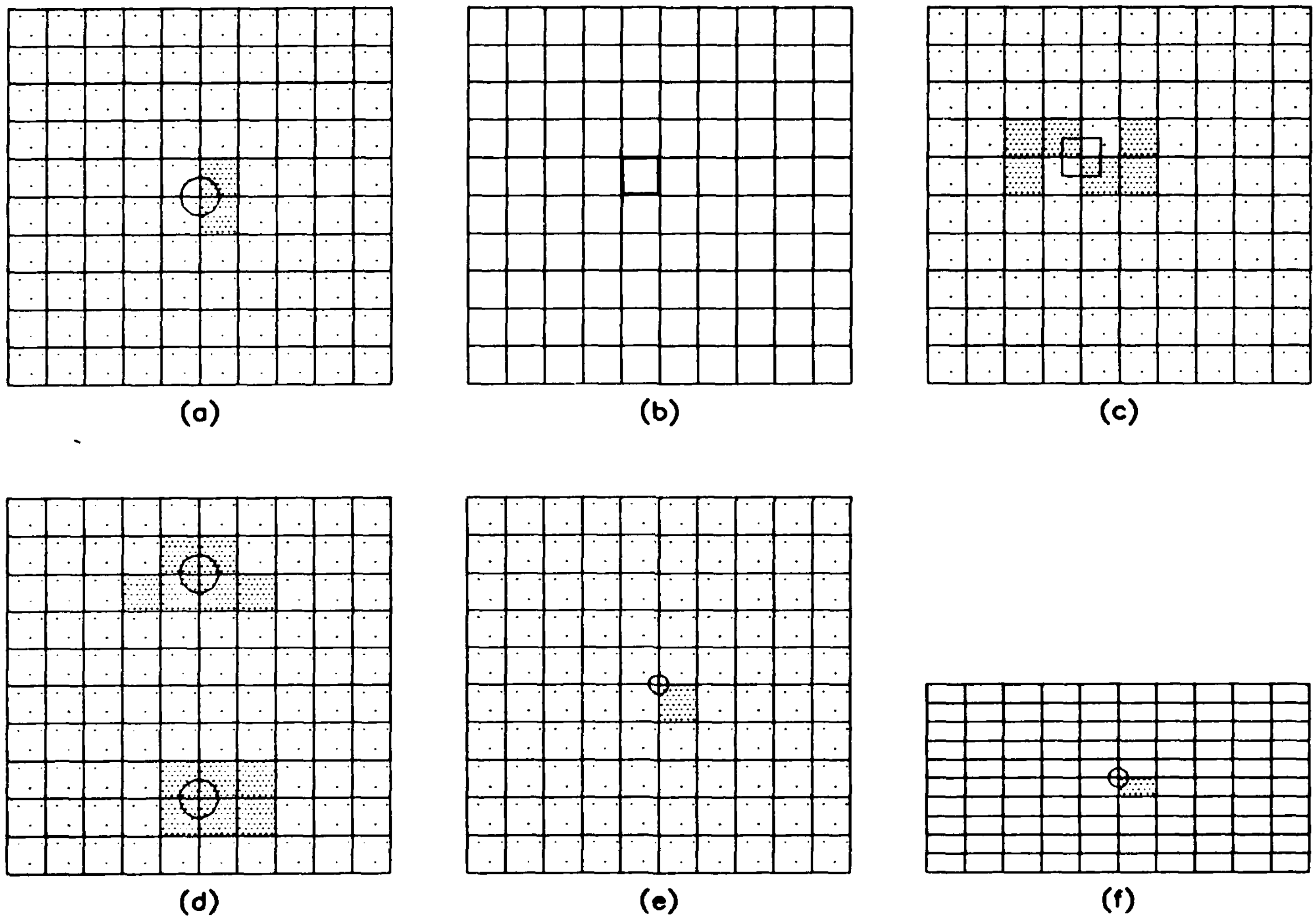


Fig. 7.19. Result of combining the CW and TOF reconstructions in Figs. 7.12 and 7.15 for the CG-GPM algorithm.

For identifying and locating tunnels none of the algorithms significantly outperformed the others.

b) Reconstructions using CW data were, in almost all cases, able to detect the anomalies in the profile; the same was not true for TOF reconstructions. However, reconstructions using TOF data, when they did detect an anomaly, were able to more accurately locate its position. This phenomenon is due to the fact that although the diffraction of the rays causes a deep attenuation in the CW measurements, it also causes the anomaly to appear larger than its actual size. On the other hand, TOF measurements are not as greatly affected by diffraction of the rays. In addition, reconstructions using TOF data were more sensitive to whether or not the anomaly 'fit' exactly on the reconstruction pixels. Again, diffraction of the rays for the CW measurements made placement of the grid less critical. This suggests the possibility of performing multiple TOF reconstructions with the grid relocated for each reconstruction.

c) Combining CW and TOF data enables one to detect anomalies and to give good insight into their composition.

7.5 Additional Reconstructions Using the CG-GPM Algorithm

7.5.1 Introduction

In this section some additional reconstructions will be performed in order to demonstrate the effectiveness of some of the methods in reducing the effects of noise in the data and reducing the effects of diffraction of the rays. For all of the reconstructions presented in this section, the CG-GPM algorithm will be used since it, in general, produced the best results for the examples given in the last section.

7.5.2 The Effectiveness of the WLS Method

The weighted least squares (WLS) method was presented in Chapter 4 as a means of adding *a priori* information to the reconstruction process. In fact, all of the reconstructions presented up until this point have used the WLS method. We would like to present some results which compare solutions obtained with and without the WLS method. For these reconstructions, the signal-to-noise ratio (SNR) will be lower (i.e., 25 and 20 dB) than what was used in the last section. The result of using these lower values of SNR is demonstrated in Figs. 7.20 and 7.21, where the electric field versus borehole depth is plotted with and without added noise, for the case of a tunnel in a homogeneous earth. The SNR was equal to 25 dB in Fig. 7.20, and equal to 20 dB in Fig. 7.21. Note how badly distorted the noisy data is in Fig. 7.21.

The reconstructions which will be presented will be for CW data generated from the two-tunnel profile, (d) in Table 7.1, and the small tunnel profile, (e) in Table 7.1. These profiles are two of the more difficult ones to reconstruct. Fig. 7.22 shows reconstructions for the two-tunnel anomaly. The images shown in (a), (b), and (c) are the results of using the least squares method, the WLS method with path length weighting, and the WLS method with path length and estimated power weighting, respectively. For a discussion of these methods refer back to Chapter IV. For these images the SNR was equal to 25 dB. Note the extraneous pixels in the image in (a) which are not present in (b) and (c). The images shown in (d), (e) and (f) are the same as those in (a) - (c), except that the SNR was equal to 20 dB. Note that the image found using the WLS method with path length and estimated power weighting gives the best indication of the presence of the tunnels.

Fig. 7.23 shows the result of attempting to detect the small tunnel. The images shown in (a), (b), and (c) use the same methods as in (a), (b), and (c) of Fig. 7.22. Again note that fewer extraneous

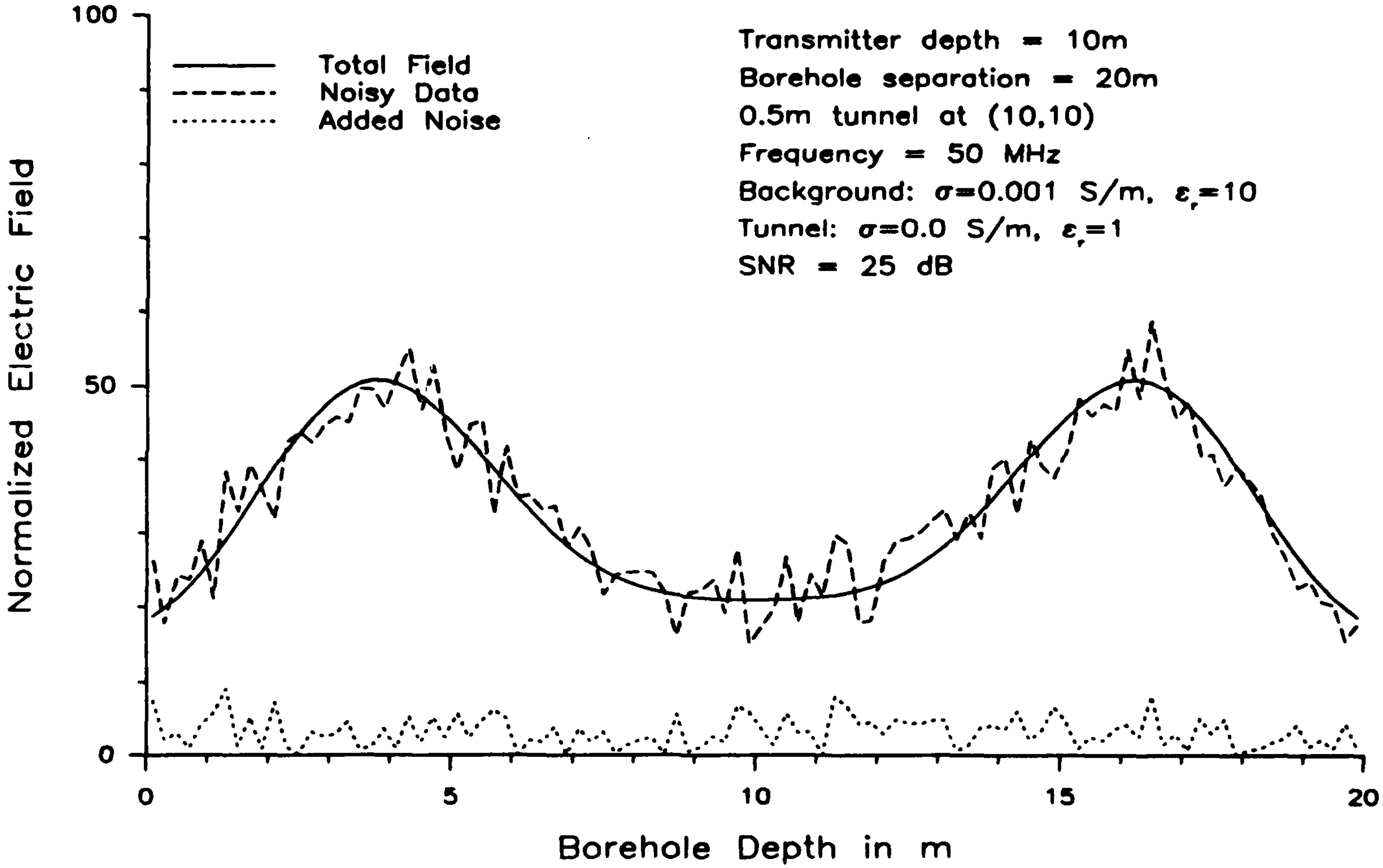


Fig. 7.20. Magnitude response of a 0.5 m tunnel in a homogeneous earth showing the effects of adding random noise, SNR = 25 dB.

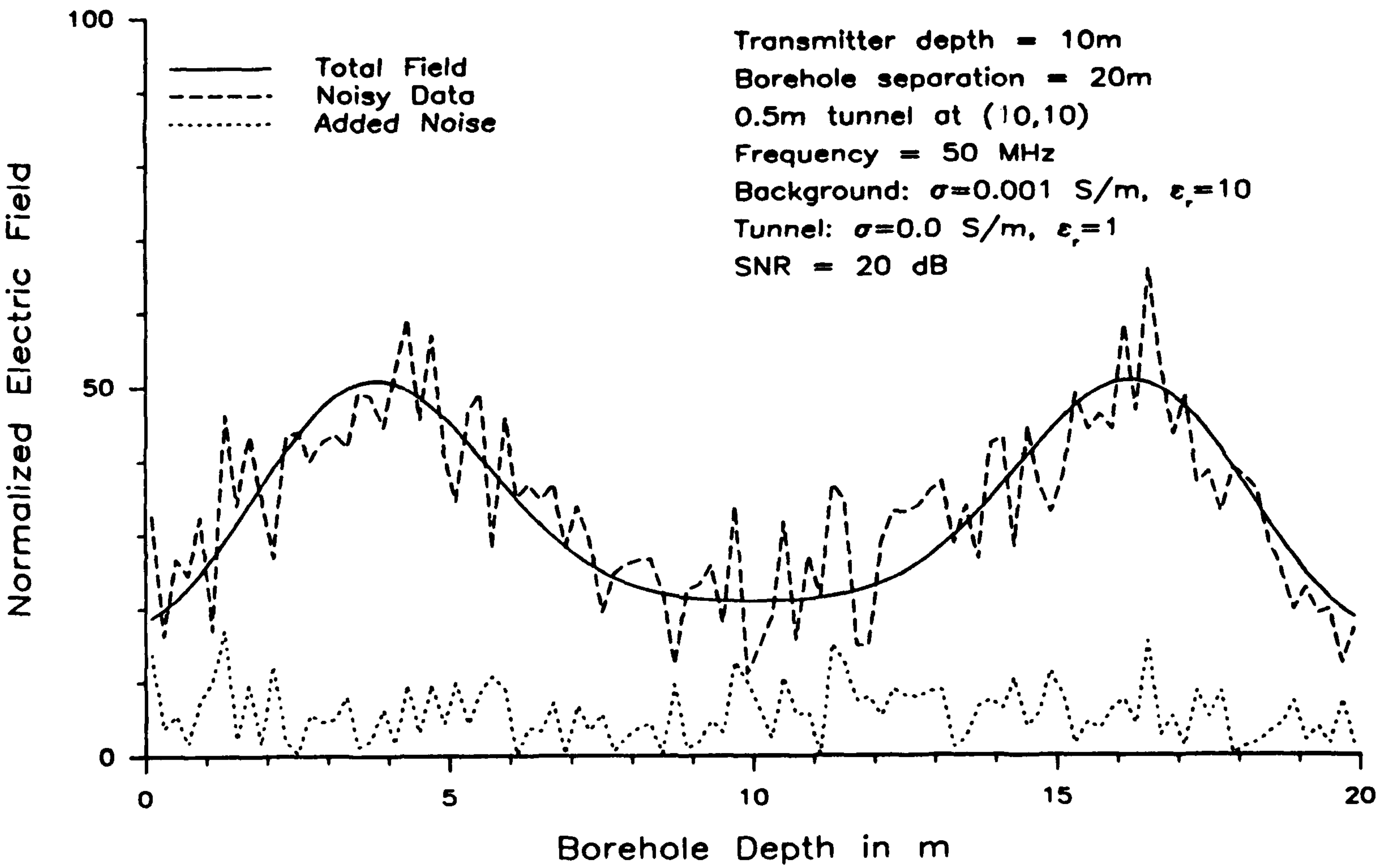


Fig. 7.21. Magnitude response of a 0.5 m tunnel in a homogeneous earth showing the effects of adding random noise, SNR = 20 dB.

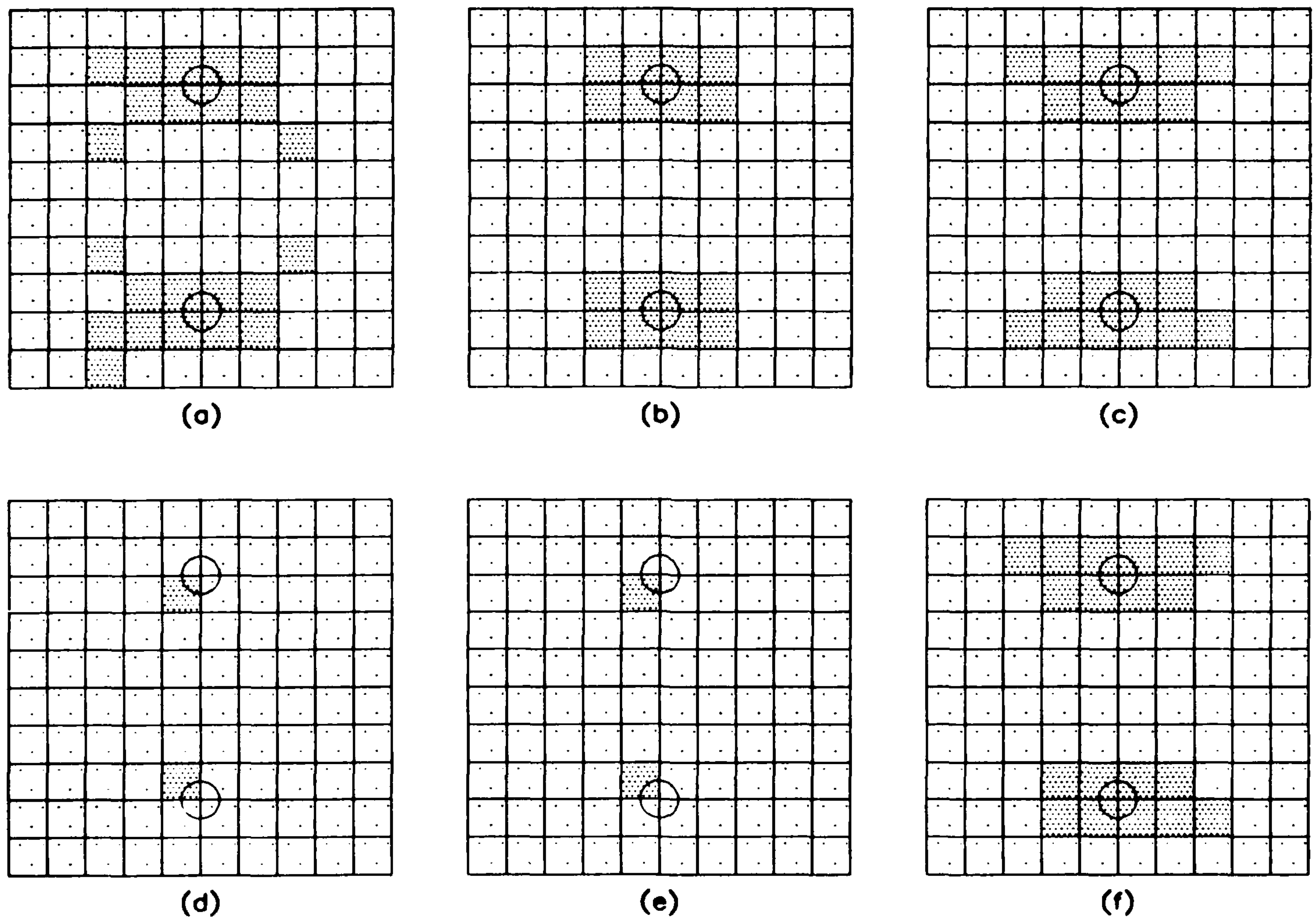


Fig. 7.22. Reconstructions for profile (d), in (a)-(c) SNR = 25 dB, in (d)-(f) SNR = 20 dB. (a) LS. (b) WLS, path weighting. (c) WLS, path and power weighting. (d) LS. (e) WLS, path weighting. (f) WLS, path and power weighting.

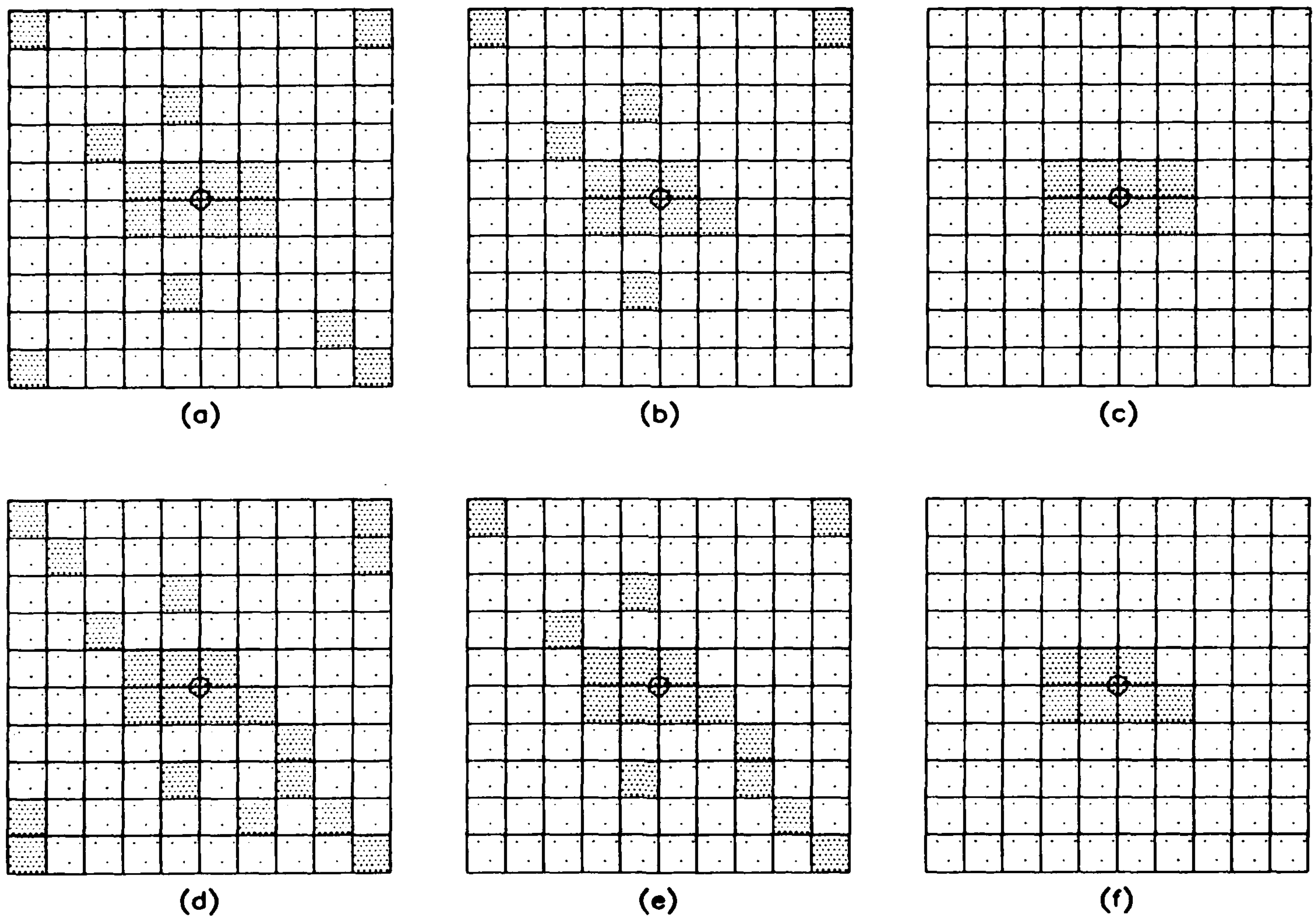


Fig. 7.23. Reconstructions for profile (e), in (a)-(c) SNR = 25 dB, in (d)-(f) SNR = 20 dB. (a) LS. (b) WLS, path weighting. (c) WLS, path and power weighting. (d) LS. (e) WLS, path weighting. (f) WLS, path and power weighting.

pixels are present in the images in (b) and (c) than in the image in (a). The images shown in (d), (e) and (f) are the same as those in (a) - (c), except that the SNR was equal to 20 dB. Again the image found using the WLS method with path length and estimated power weighting gives the best indication of the presence of the small tunnel.

As these examples show, the WLS method is extremely effective for reconstructing images under conditions in which the data is very noisy. Since it is expected that actual measurements will contain considerable noise, the WLS method should be used in all cases.

7.5.3 The Effectiveness of Constraining the Solution Image

In this section we present results which show the effectiveness of applying constraints to the reconstructed image. Recall that the results in Section 7.4 were obtained by constraining the pixel values in the reconstructed images to be greater than zero. The discussion in Chapter III gives reasons for considering constrained solutions. In fact, the CG-GPM algorithm was developed in Chapter IV as a means of finding a constrained solution. We again consider detection of the two tunnels and the small tunnel in a homogeneous earth using CW measurement data. Signal-to-noise ratios of 25 and 20 dB will be used.

As already stated, the background medium has a conductivity equal to 0.001 S/m and a relative permittivity equal to 10. These values result in an attenuation equal to 0.06 Np/m at 50 MHz. When the solution is constrained to this value of attenuation, the resulting image has all pixel values equal to this value of attenuation. Instead, we shall use constraint values equal to 0.00, 0.01, 0.02, and 0.03 Np/m.

The results of using the above constraint values for the two-tunnel profile are shown in Fig. 7.24, with (a) having a constraint value of zero and (d) having the highest constraint value of 0.03 Np/m. (b) and (c) have the intermediate constraint values. Note that the

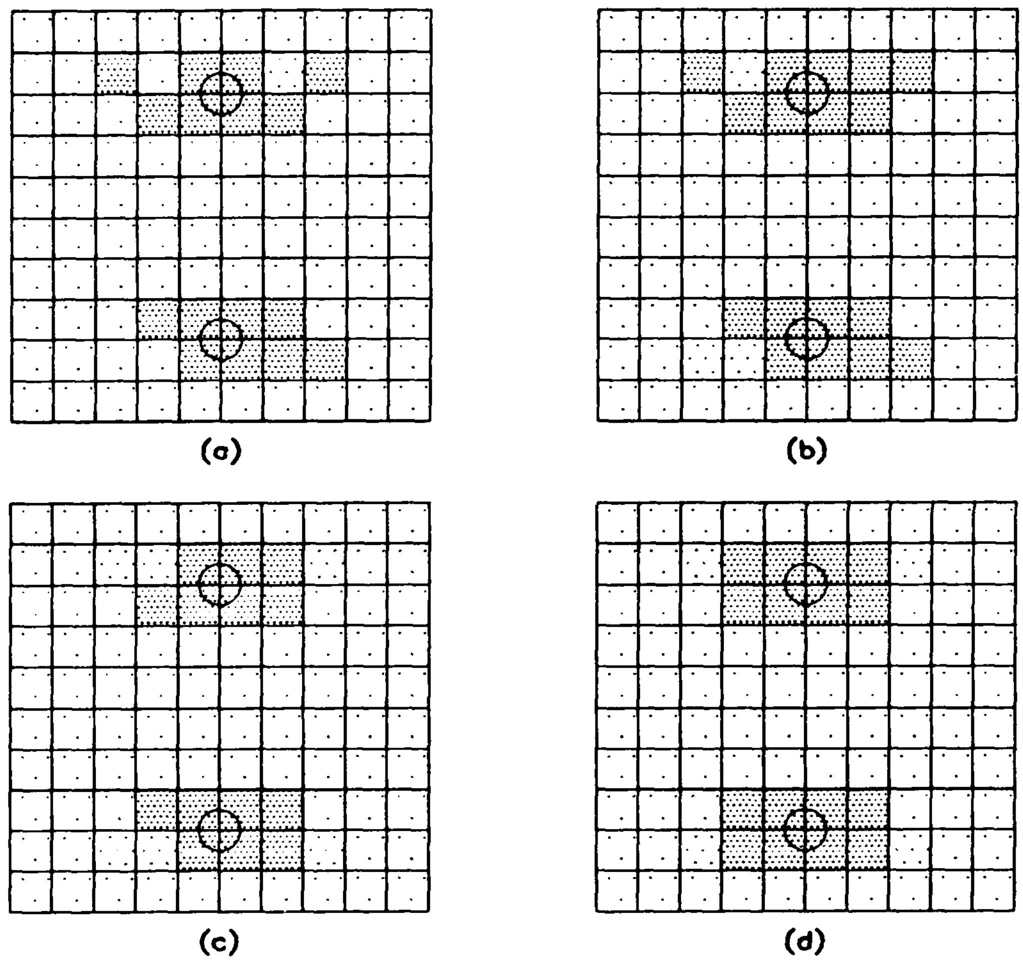


Fig. 7.24. Reconstructions for profile (d), SNR = 25 dB, constraint values as follows: (a) 0 Np/m. (b) 0.01 Np/m. (c) 0.02 Np/m. (d) 0.03 Np/m.

image in (d) best identifies the anomalies in that the extraneous pixels are localized adjacent to the anomalies' positions. The same constraint values were applied to the reconstruction of the small cylinder, and the results are shown in Fig. 7.25. The image in (a) of the figure has regions of low attenuation directly above and below the anomaly. These low attenuation regions are artifacts of the diffraction effects which cause the peaks in the electric field at borehole depths of 4 and 16 m in Figs. 7.20 and 7.21. Note that the artifacts have been eliminated in (c) and (d) of Fig. 7.25.

With the SNR equal to 20 dB, the two tunnels cannot be identified in the images (a) and (b) of Fig. 7.26. However, for constraint values of 0.02 and 0.03 Np/m (c and d of the figure) the anomalies are clearly distinguishable. Fig. 7.27 shows the reconstructions for the small tunnel with the SNR equal to 20 dB. Again, the best images are for constraint values equal to 0.02 and 0.03 Np/m. Note also that these reconstructions are superior to the one in Fig 7.13 (e), where the SNR was equal to 30 dB. In that figure the constraint value was equal to zero.

In summary, by using constraints the reconstructions are less susceptible to the effects of noise as demonstrated in Figs. 7.24 - 7.27. In addition, the application of constraints reduces the unwanted artifacts due to the diffraction of the electromagnetic waves. In general the constraint value should be chosen to be less than the attenuation value of the background medium. For the cases investigated above, one third of the background attenuation value was a good compromise between reducing unwanted artifacts, and washing out the image entirely.

7.6 Ray Optics Refinement

7.6.1 Introduction

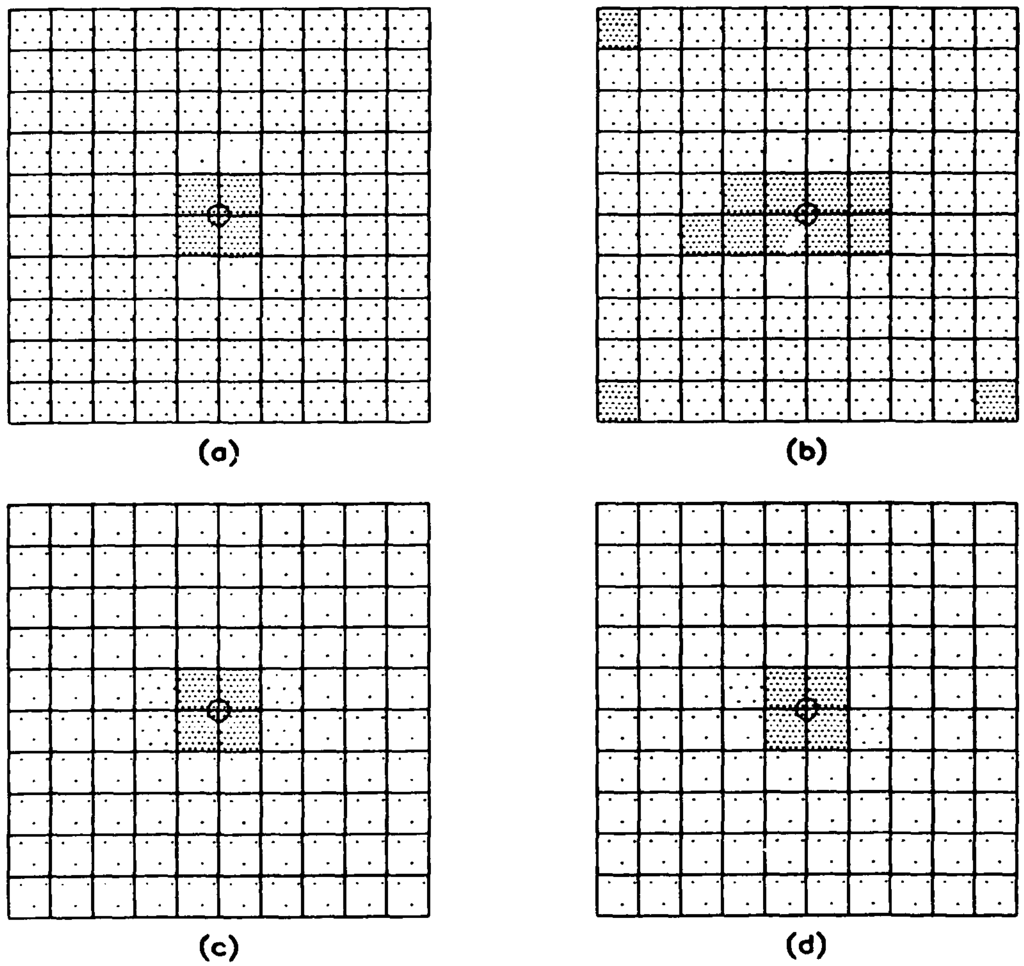


Fig. 7.25. Reconstructions for profile (e), SNR = 25 dB, constraint values as follows: (a) 0 Np/m. (b) 0.01 Np/m. (c) 0.02 Np/m. (d) 0.03 Np/m.

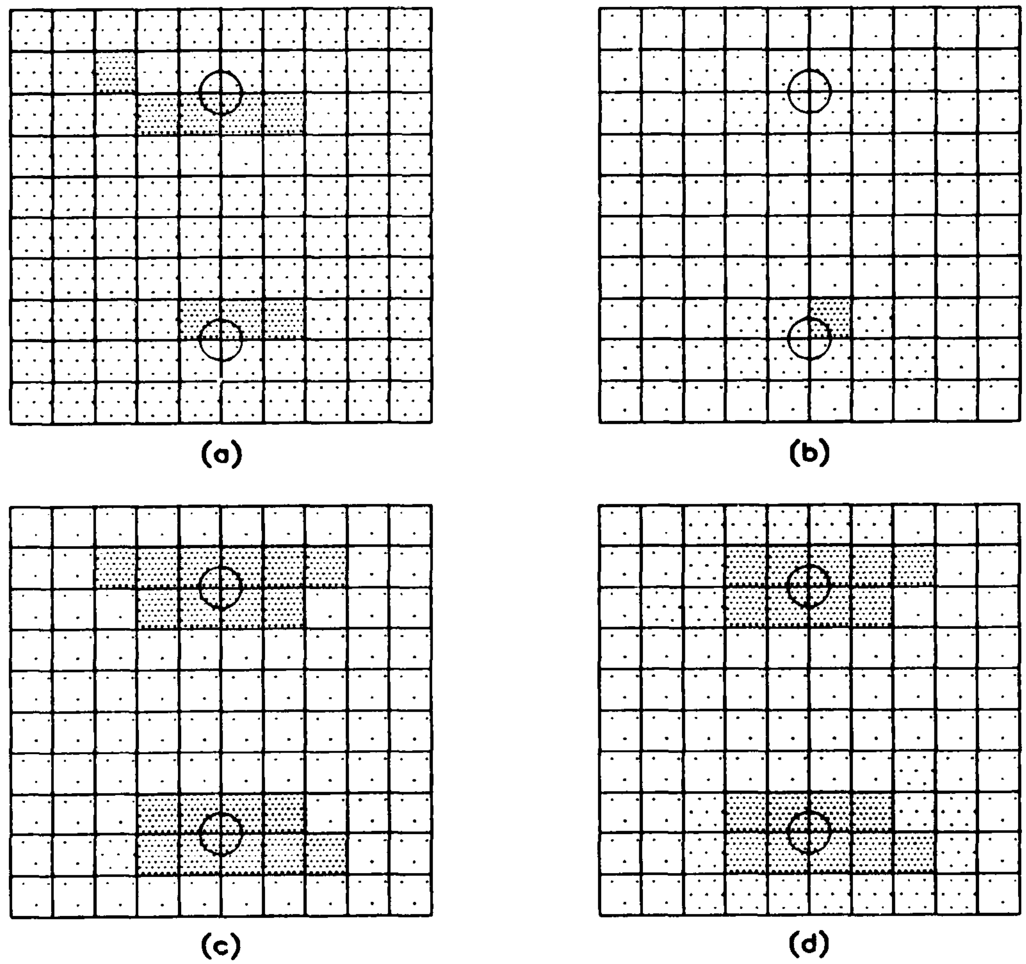


Fig. 7.26. Reconstructions for profile (d), SNR = 20 dB, constraint values as follows: (a) 0 Np/m. (b) 0.01 Np/m. (c) 0.02 Np/m. (d) 0.03 Np/m.

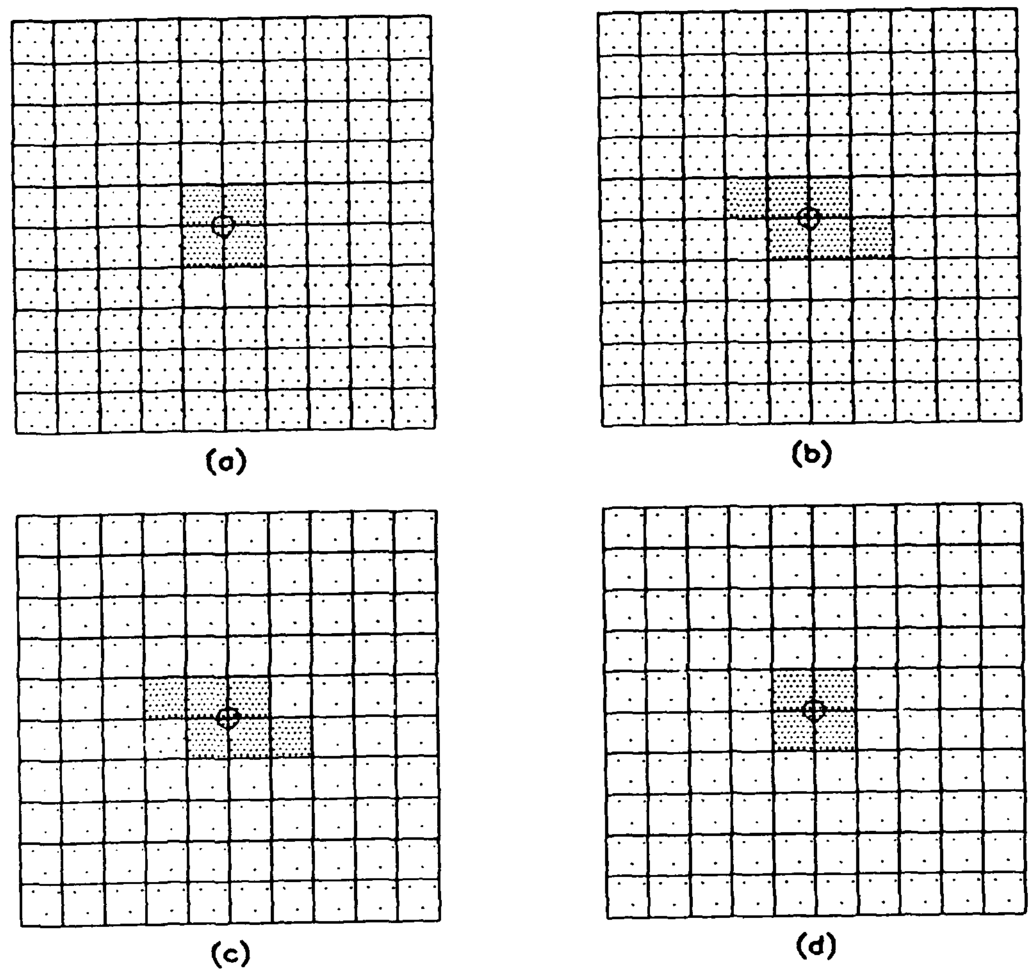


Fig. 7.27. Reconstructions for profile (e), SNR = 20 dB, *constraint values as follows:* (a) 0 Np/m. (b) 0.01 Np/m. (c) 0.02 Np/m. (d) 0.03 Np/m.

In this section we will present methods for improving reconstructed images. As can be seen from the examples presented previously, the reconstructed images are not ideal in that the images do not exactly match the actual cross section of the earth being investigated (i.e., the reconstructions are not perfect). Reasons for this mismatch include:

a) The straight ray model, which is used as the basis for the algebraic inversion method, neglects the reflections, refractions, and diffractions of the rays.

b) The finite resolution of the pixels does not allow the adequate representation of some images.

Note that the shortcoming discussed in b cannot always be resolved by decreasing the size of the pixels, since this will result in a (computationally) longer and more instable inversion process. Two methods for accounting for non-straight rays are ray tracing and the ray optics model matching procedure discussed in Section 3.2.3.

Ray tracing is a method for improving a reconstructed image by finding the paths rays would take through the current estimate of the cross section in order to obtain a better indication of the actual ray paths [14], [15], [66]. This new path length information is then used to construct a new coefficient matrix for (3-26), and then this equation can be inverted using the methods of Chapter IV to find an improved image. The ray tracing process can be repeatedly applied in order to obtain better estimates of the ray paths and therefore better reconstructed images. This method normally considers the reflection and refraction of the rays at interfaces between pixels, and ignores diffraction effects. The refraction and reflection angles can be calculated from Snell's law [66]. However, from the discussion in Section 2.5.2, Snell's law will not always be a good approximation for the refraction of rays in lossy media.

The other method for refining the reconstructed image was described in Section 3.2.3. The main assumptions in using this method

are that the region contains high contrast anomalies and that diffraction effects are dominant (over refraction effects). The reconstructed image will be used to locate regions where a high contrast anomaly might exist and then the model matching method will be used to locate the anomaly within the region(s). The second assumption is necessary since a theory for adding the contributions from refracted and diffracted rays in the reconstruction process does not presently exist.

From the discussion above, we can conclude that ray tracing will be effective when refraction is the dominant effect, and the model matching procedure will be effective when diffraction is the dominant effect. As noted in Chapter II, diffraction will be significant when the index of refraction of the scatterer (anomaly) differs from the index of refraction of the background. If we consider as an example a background earth having a conductivity equal to 0.001 S/m and a relative permittivity equal to 10, then it can be seen from Fig. 2.26 that if the conductivity of the anomaly differs from the background but its permittivity is the same, then the index of refraction of the anomaly will be comparable to the background, and diffraction will not be significant. However, if the permittivity of the anomaly differs from the background, then the index of refraction of the anomaly could be greatly different from the background (Fig. 2.27) and diffraction will be significant. In most cases, a region having a higher permittivity will also have a higher conductivity [24], so that a technique based on diffraction theory will be the best choice.

7.6.2 Ray Tracing Using Snell's Law

A. Introduction

The method of using ray tracing was described above, and a more complete description is given in [66]. See [84] for details of the

algorithmic process for finding the ray paths between transmitters and receivers. Note that in [84] Snell's law was used for finding the direction of travel of the refracted rays. However, we have shown in Chapter II that in lossy media an extension of Snell's law must be used. In particular, see (2-54) through (2-57) for a description of the transmission of rays through an interface between lossy media. Note that in tracing a refracted ray, the ray corresponding to the direction of constant phase is the one being considered. Refer again to Fig. 2.22 for clarification of these ideas. The interfaces that are being considered are the boundaries between adjacent pixels in a reconstructed image.

Fig. 7.28 gives an illustration of the process of finding the rays linking a single transmitter to a number of receiver locations. The rectangle in the center of the figure represents a region having lower conductivity (and permittivity) than the background. Some of the items of interest in this figure include:

1) The ray tracing algorithm could not find ray paths linking some of the receivers to the transmitter. This failure of the algorithm is due the discontinuity caused by the corners of the rectangle.
2) The presence of the rectangle causes the ray paths to deviate from the straight line connecting the transmitter to the receivers. In particular, if many rectangles were present (as would be the case in a reconstructed image where each pixel is a rectangle) then the paths would be very erratic.

Physically, one does not expect that the rays would behave as described above in an actual cross section of the earth where abrupt changes in the medium do not normally occur. The imposition of a discrete nature to the cross section (via the pixels) will cause the estimated ray paths to not coincide with the actual ray paths. It is for this reason that ray tracing cannot be applied to the unprocessed reconstructed image.

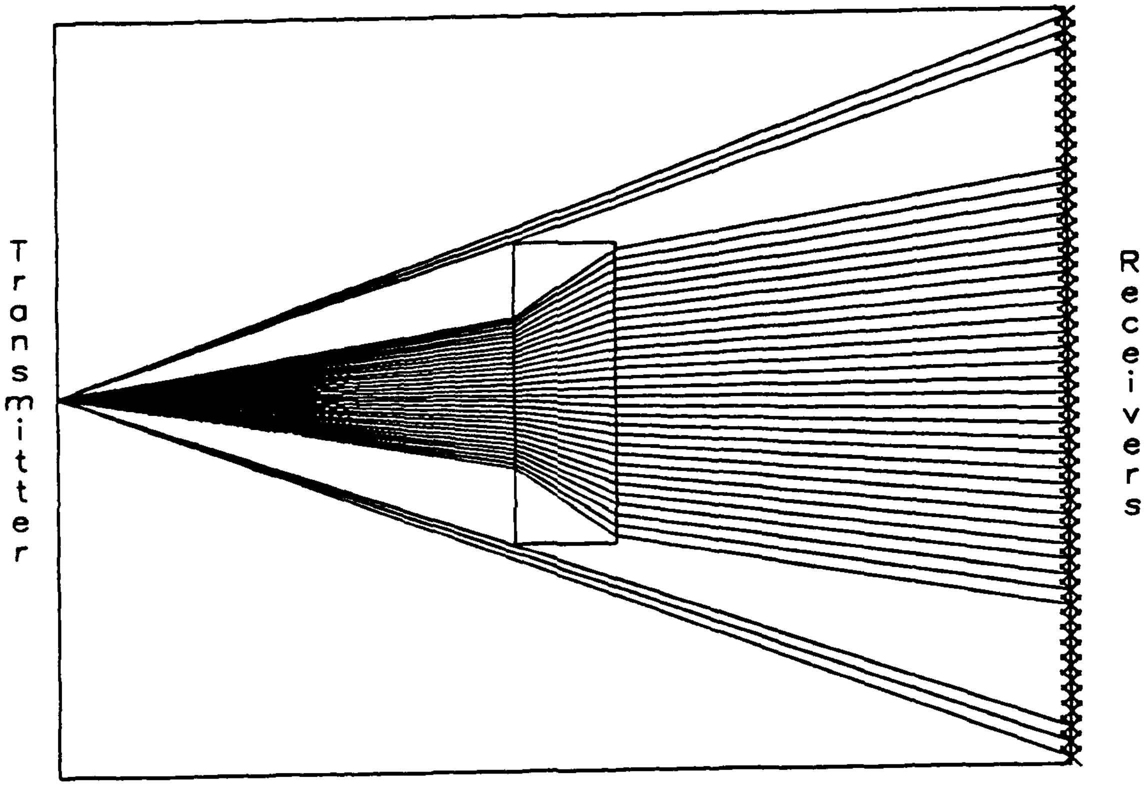


Fig. 7.28. Illustration of rays refracting through a low conductivity rectangular cylinder.

It was suggested in [66] that filtering of the type described in Chapter V be applied to the reconstructed image in order to reduce the sharp discontinuities between pixels. We have also found that it helps to partition the image (using the MMP, for example) prior to tracing rays. This partitioning in effect removes the 'noise' in the background image. This is important since even if two pixels have similar (but not identical) electrical parameters, a ray which exits the first pixel and enters the second pixel at an angle far from normal will experience a large refracting angle. This phenomenon can be seen in Fig. 7.28 where the rays which enter the rectangle at larger angles from normal are refracted more.

Another technique for ensuring that the traced rays more closely represent the actual ray paths taken in the measurement process is to use some sort of interpolation between pixels in the reconstructed image [14], [15]. This interpolation will also remove some of the adverse effects of discretizing the cross section into pixels. We have found that a linear interpolation [15] gives the best results at the least computational expense. In judging the merit of an interpolation scheme, we consider the scheme which is able to find the most rays linking transmitter and receiver locations to be the best. In using linear interpolation, if a ray enters at the center of a pixel, then we use the parameters of that pixel and the exiting pixel in (2-54) through (2-57) in determining the refracted ray. If, however, the ray enters below the center of the pixel, then a linear combination of that pixel and the one below it (as well as for the exiting pixel) are used in (2-54) through (2-57). Fig. 7.29 shows an illustration of tracing a ray through an interface between two pixels. The electrical parameter of the entering pixel for this example would be

$$X = (1 - d/H)\, X_1 + (d/H)\, X_2 , \tag{7-2}$$

where d is the distance from the center of the pixel to the point

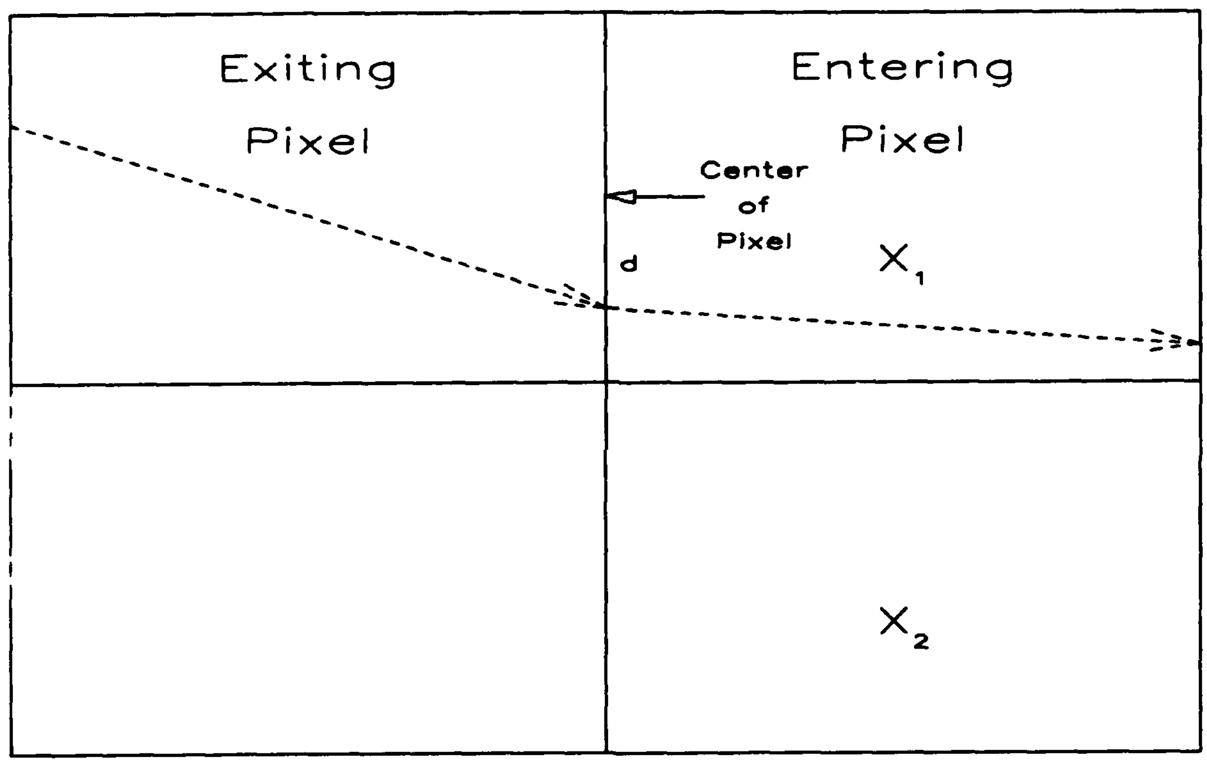


Fig. 7.29. Illustration of using interpolation in the ray tracing process.

where the ray enters the pixel, H is the height of the pixel and X_1 (X_2) is the electrical parameter of the upper (lower) pixel. The same process is applied to find the electrical parameters of the pixel which the ray exited.

Figs. 7.30 and 7.31 illustrate the usefulness of the techniques described above. Fig. 7.30 shows an example of ray tracing through a raw reconstructed image. In this figure six transmitter and receiver locations are shown. Note that the ray trace algorithm is unable to find paths linking all the transmitters and receivers, and the paths which are found are very erratic. Fig. 7.31 shows the result of tracing rays using the techniques described above. Note that in this figure, all the paths linking transmitters and receivers have been found. In addition, the paths are not as erratic as those found in Fig. 7.30. We will first apply the ray tracing process to reconstructions using CW data, and then to reconstructions using TOF data.

B. Ray Tracing for CW Data

In this section we will show examples of using ray tracing to improve reconstructions obtained using CW data. As described above, ray tracing will only be effective when refraction effects are more dominant than diffraction effects. Therefore, we will only consider profiles where this criterion is met.

The ray tracing process was applied to the reconstruction of a high conductivity circular cylinder in a homogeneous earth. The background earth has a conductivity of 0.001 S/m and a relative permittivity of 10. The cylinder has a conductivity of 0.004 S/m and a relative permittivity of 10. Since the permittivity of the background and the cylinder are the same, we expect refraction will be the dominant effect. Fig. 7.32 is a three dimensional representation of the reconstructed image. The image was reconstructed using the

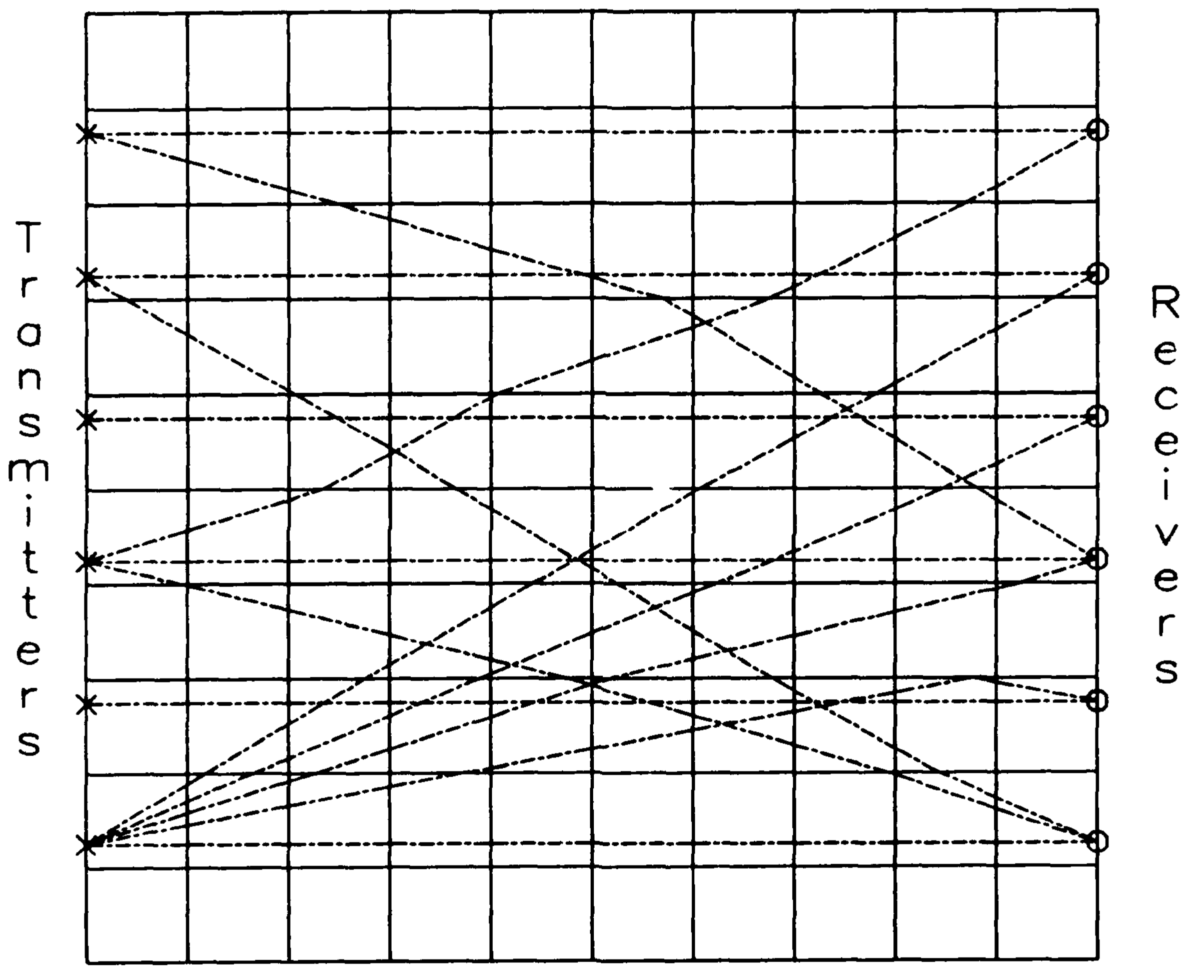


Fig. 7.30. Illustration of ray tracing process for a raw reconstructed image.

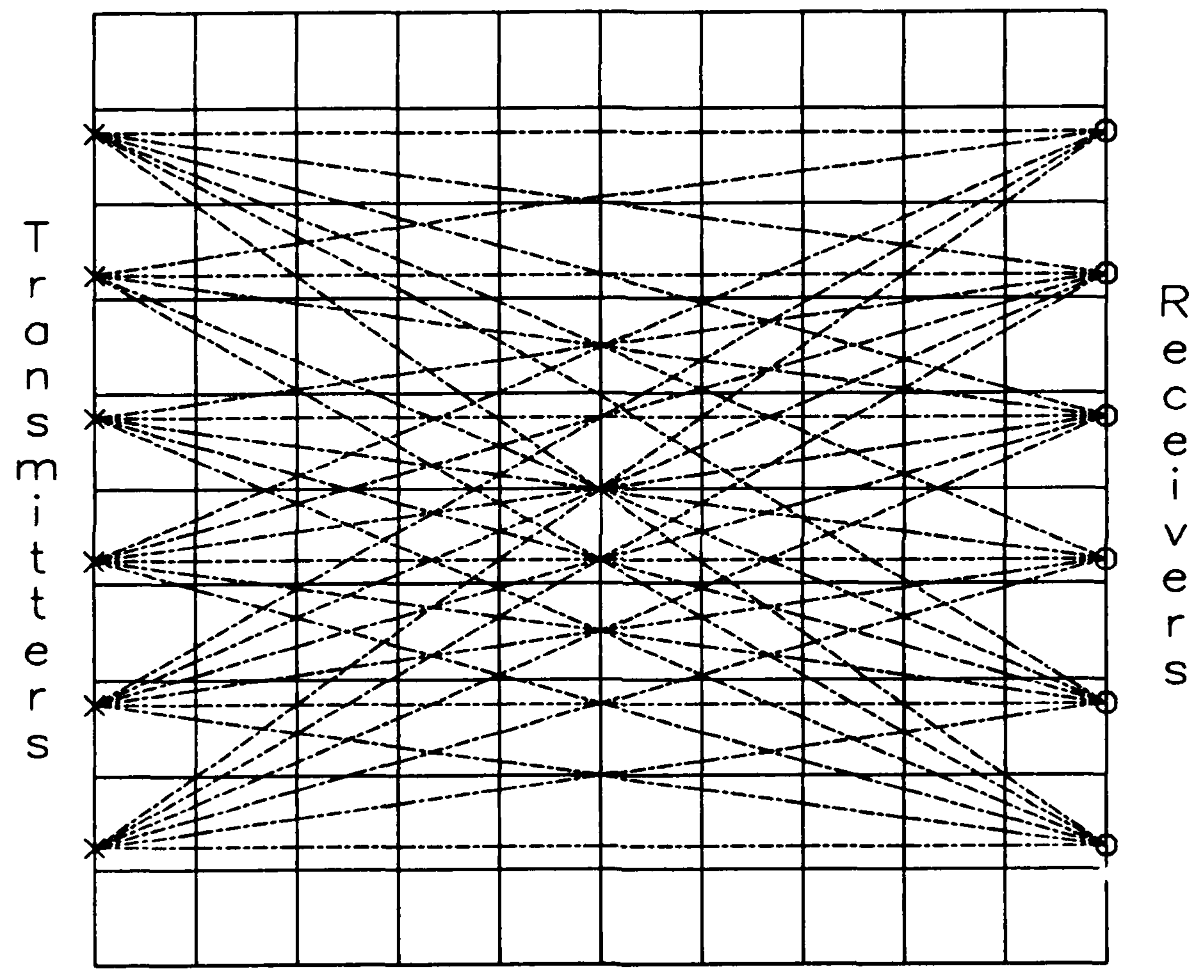


Fig. 7.31. Illustration of ray tracing process for a filtered partitioned reconstructed image using linear interpolation.

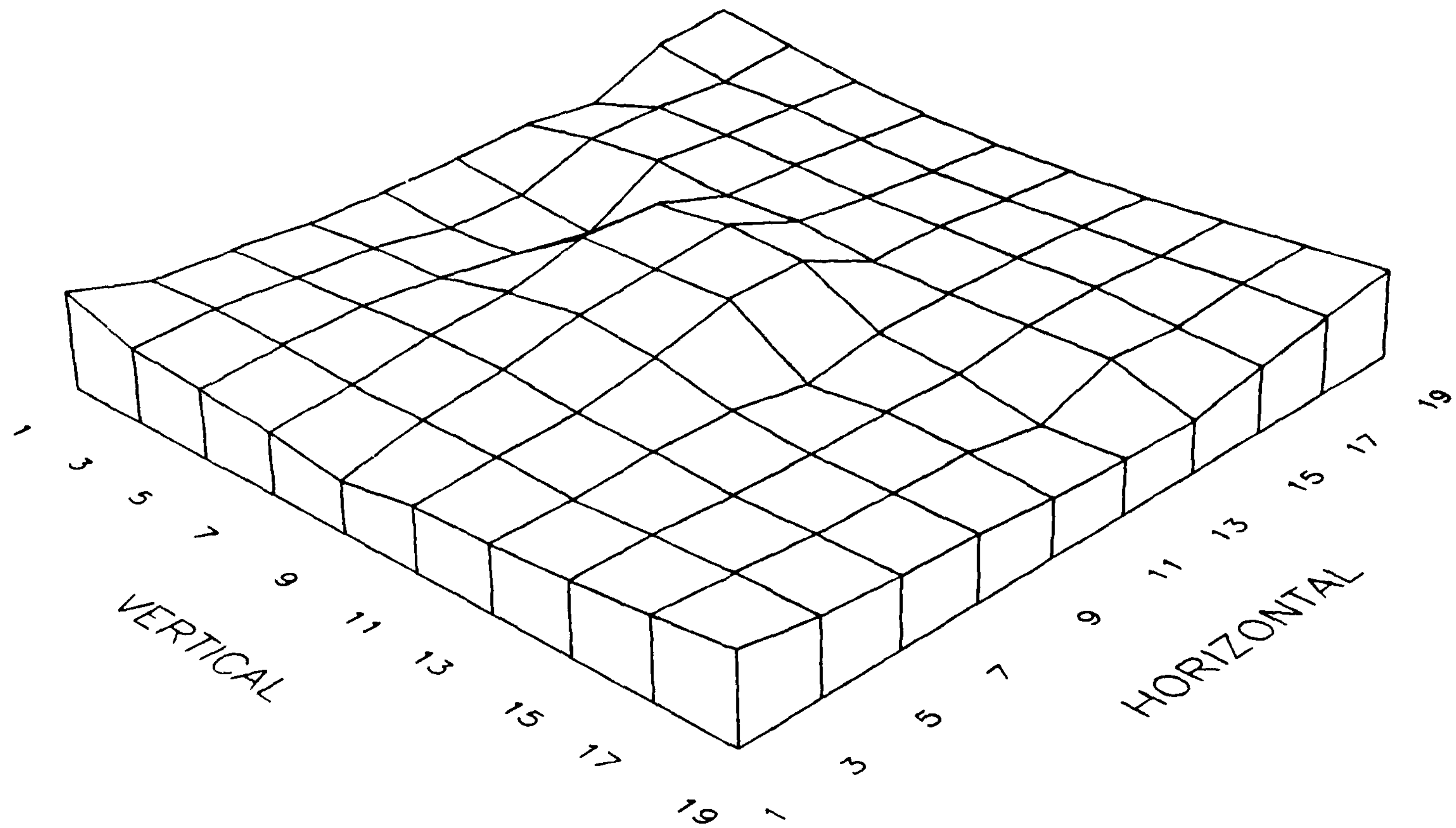


Fig. 7.32. Reconstruction for high conductivity cylinder using CW measurements.

CG-GPM algorithm with a constraint value of 0.002 S/m. Regions of higher attenuation are indicated by greater height in the figure. Note that the cylinder's presence is seen near the center of the figure, but it is not clearly distinguishable. Fig. 7.33 shows the result of applying one iteration of ray tracing to the reconstructed image (further iterations caused the solution to diverge). Note that ray tracing has caused the presence of the cylinder to be more clearly indicated, and the region surrounding the cylinder to have fewer perturbations.

Fig. 7.34 summarizes the ray tracing example using two-dimensional images. In part (a) of the figure the filtered reconstructed image is shown. Part (b) shows the result of partitioning this image into two groups using the MMP. Note the spreading of the cylinder in both horizontal and vertical directions. Part (c) is the filtered reconstructed image after applying ray tracing, with (d) showing the partitioned version of this image. Note that the pixel adjacent to the cylinder is identified as the high attenuation region. This gives a more accurate indication of the size of the cylinder although its location is in error.

C. *Ray Tracing for TOF Data*

In this section we will use ray tracing to improve the reconstructed images obtained from TOF data. Note that for TOF measurements, refracted rays will in general be more important than diffracted rays since in TOF measurements the time of the first pulse reaching the receiver will be measured. For example, Fig. 2.21 shows the pulse waveform which would be observed at the receiver, and the pulse due to the diffracted ray is the last pulse to be observed. Therfore, we can use ray tracing on the high contrast profiles considered in Section 7.4. The anomalies will either have a relative permittivity of 20 (high permittivity) or a relative permittivity of 1

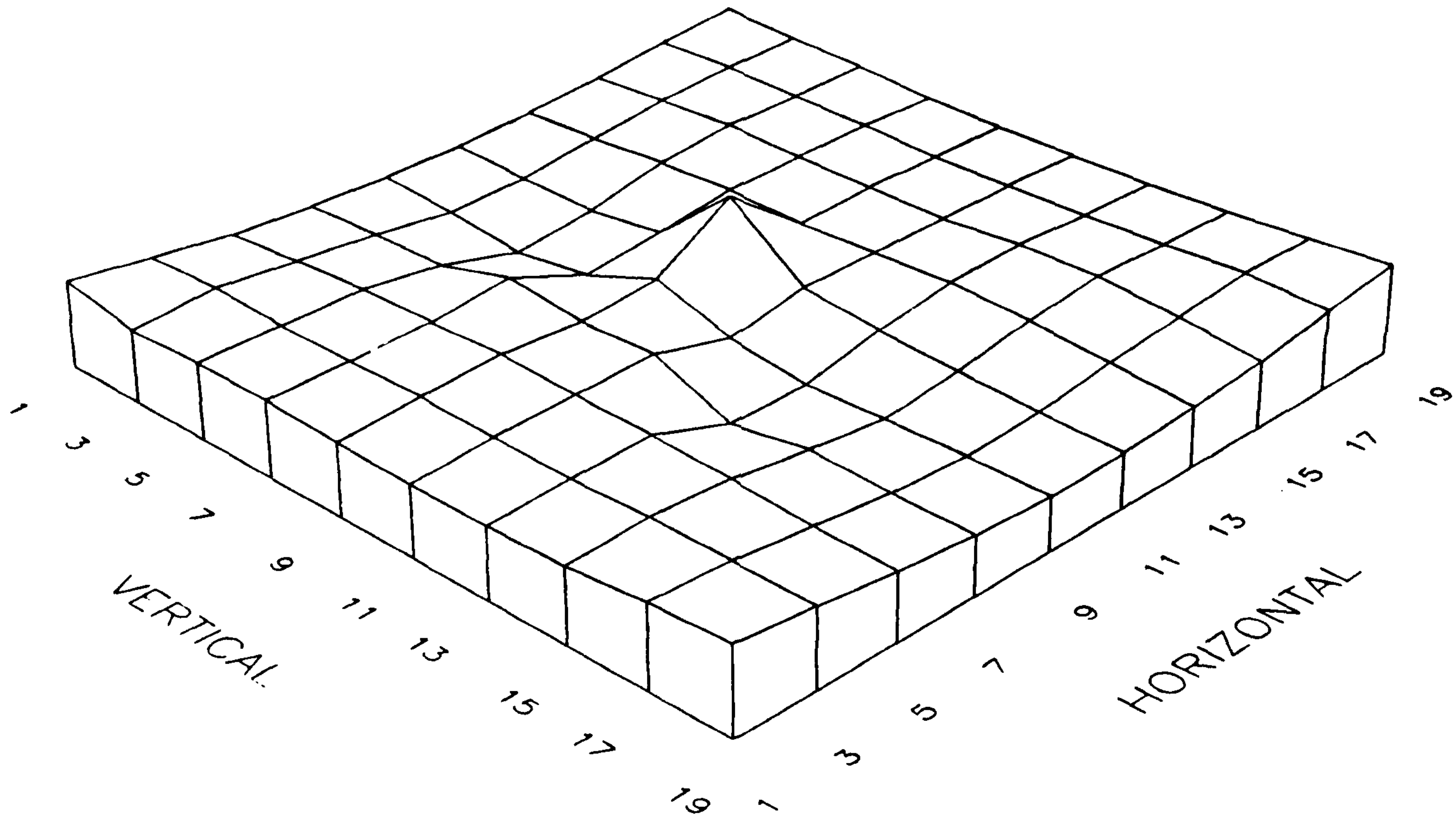


Fig. 7.33. Result of using ray tracing to improve the reconstruction in Fig. 7.31.

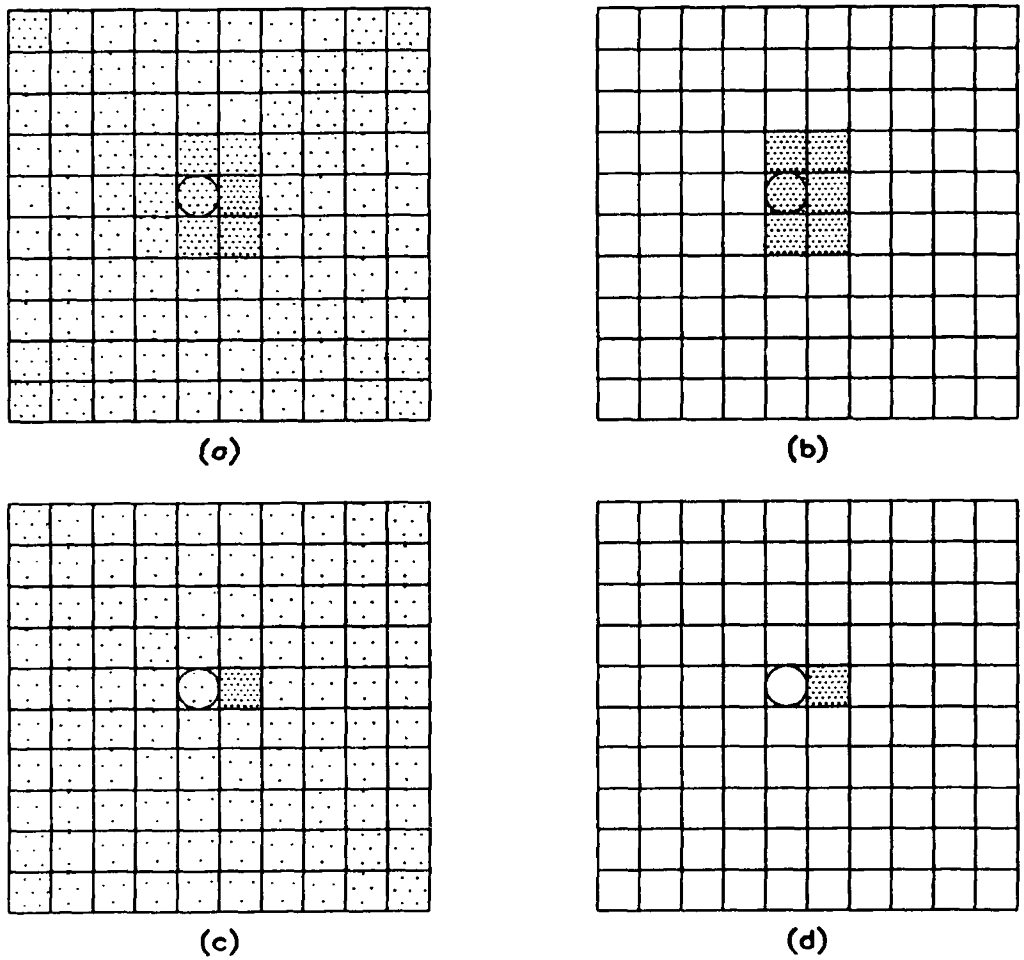


Fig. 7.34. Summary of ray tracing example. (a) Filtered reconstructed image. (b) MMP of image in (a). (c) Image resulting from ray tracing. (d) MMP of image in (c).

(tunnel). As before, the background has a relative permittivity of 10 and a conductivity of 0.001 S/m.

We first consider the profile having the square cylinder coinciding with one of the reconstructed pixels (profile b in Table 7.1). Ray tracing should be most effective on this profile since the reconstructed image will have the cylinder exactly on one of the pixels. Fig. 7.35 summarizes the results for the high permittivity cylinder. In part (a) of the figure is the filtered reconstructed image while part (b) is the result of partitioning this image into two groups using the MMP. Parts (c) and (d) of the figure are the results of using one iteration of ray tracing. As can be seen, ray tracing has not improved the image. This result is not surprising in that the reconstructed image was already a good representative of the actual cross section. Fig. 7.36 shows similar results for the low permittivity cylinder. Again, ray tracing was unable to improve a good reconstruction.

We now consider ray tracing applied to the square cylinder which does not coincide with one of the pixels (profile c in Table 7.1). Fig. 7.37 shows the reconstructed images for the high permittivity cylinder. This time ray tracing has slightly degraded the image in that the partitioned ray traced image, (d), does not indicate the presence of the cylinder. However, the filtered ray traced image, (c), does show some indication of the presence of the cylinder. For the low permittivity cylinder, the results are shown in Fig. 7.38. This time ray tracing has greatly improved the image in that the partitioned image clearly indicates the presence of the cylinder, although its location is slighlty in error.

D. Conclusions

In this section it was shown that ray tracing can be an effective method for improving reconstructed images resulting from measurements

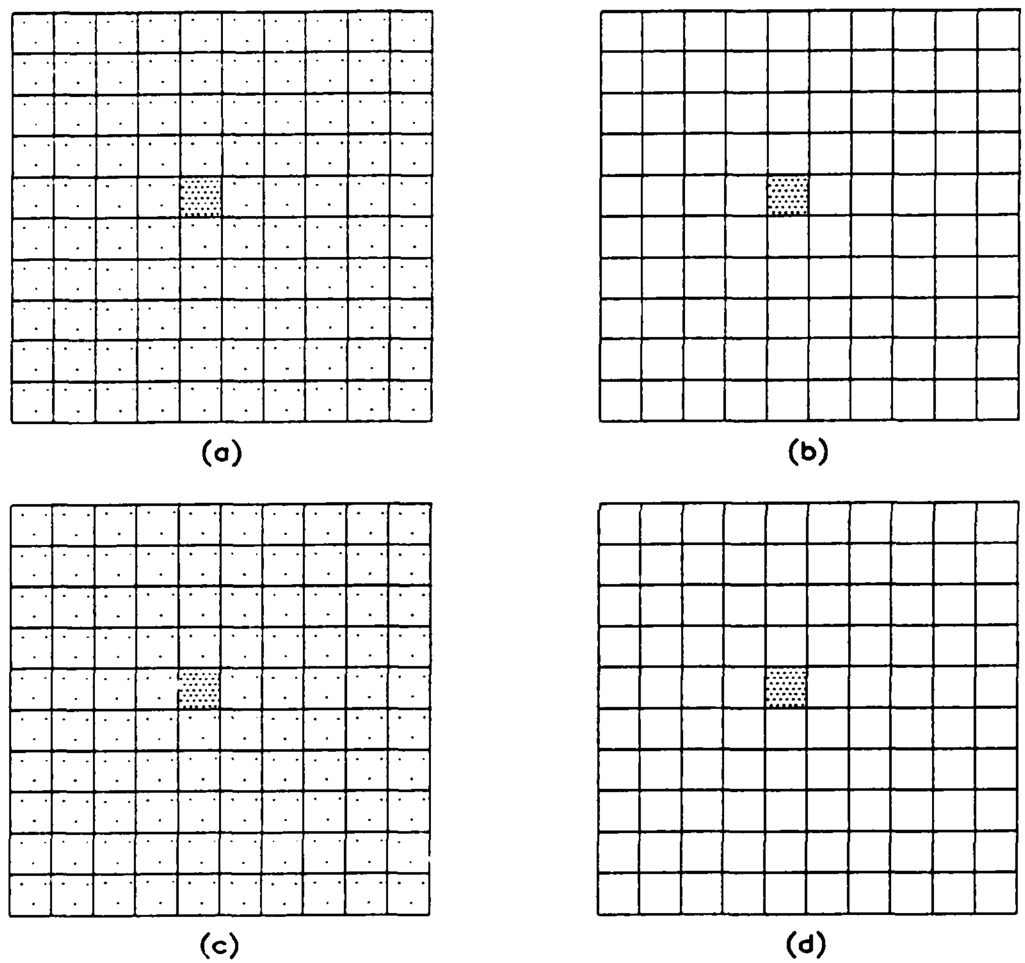


Fig. 7.35. Ray tracing results for high permittivity cylinder using TOF measurements. (a) Filtered reconstructed image. (b) MMP of image in (a). (c) Filtered ray traced image. (d) MMP of image in (c).

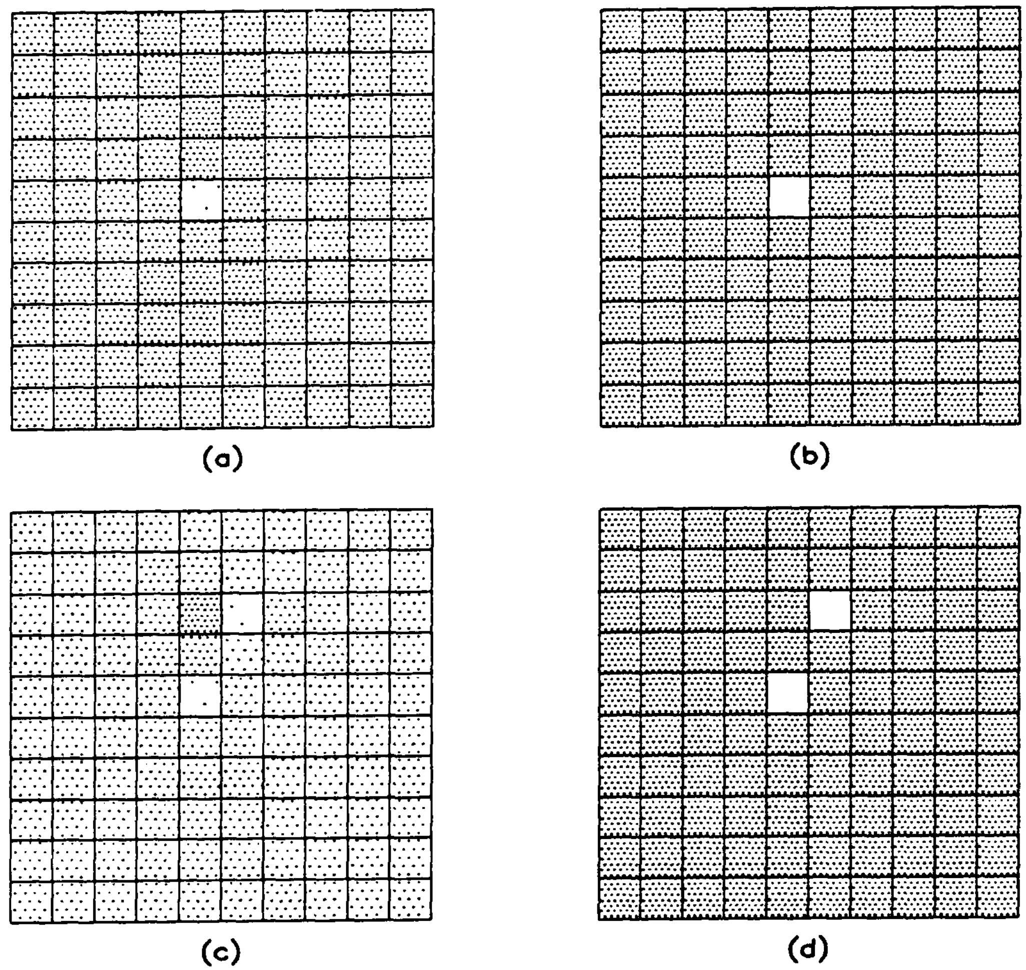


Fig. 7.36. Ray tracing results for low permittivity cylinder using TOF measurements. (a) Filtered reconstructed image. (b) MMP of image in (a). (c) Filtered ray traced image. (d) MMP of image in (c).

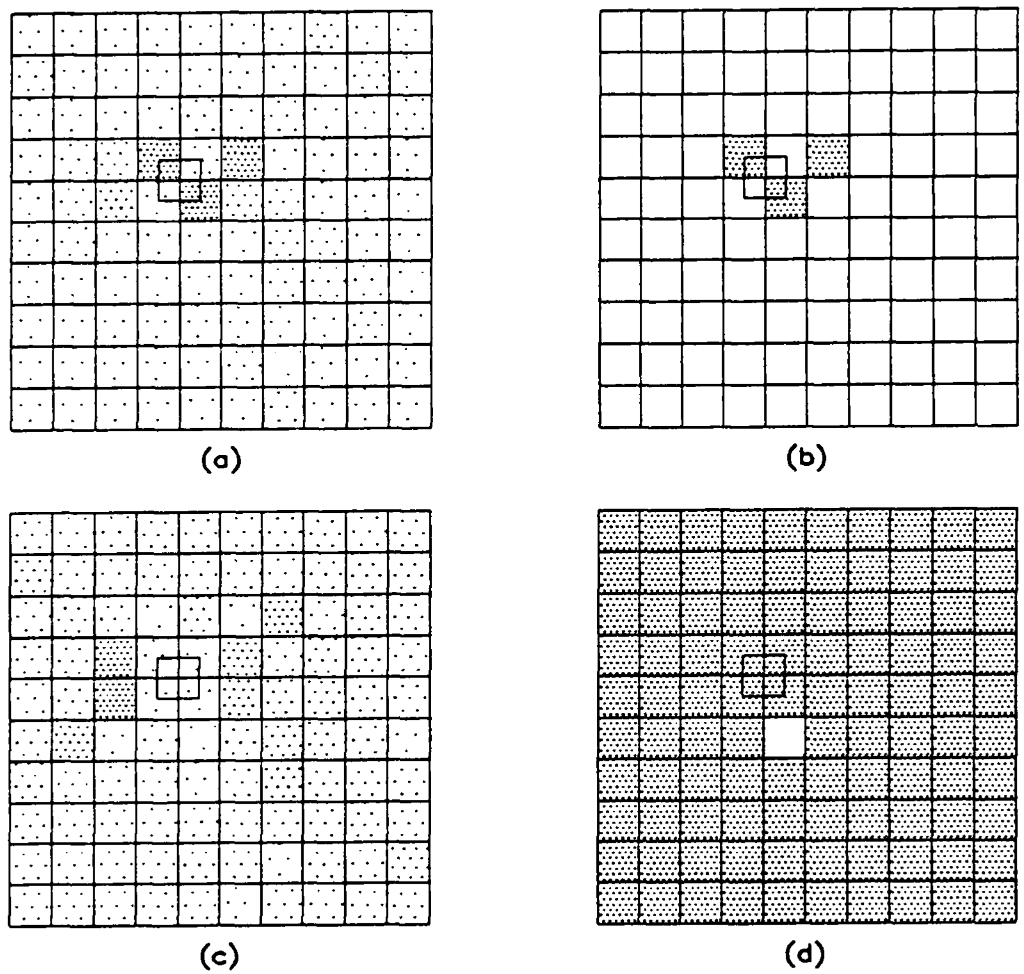


Fig. 7.37. Ray tracing results for high permittivity cylinder using TOF measurements. (a) Filtered reconstructed image. (b) MMP of image in (a). (c) Filtered ray traced image. (d) MMP of image in (c).

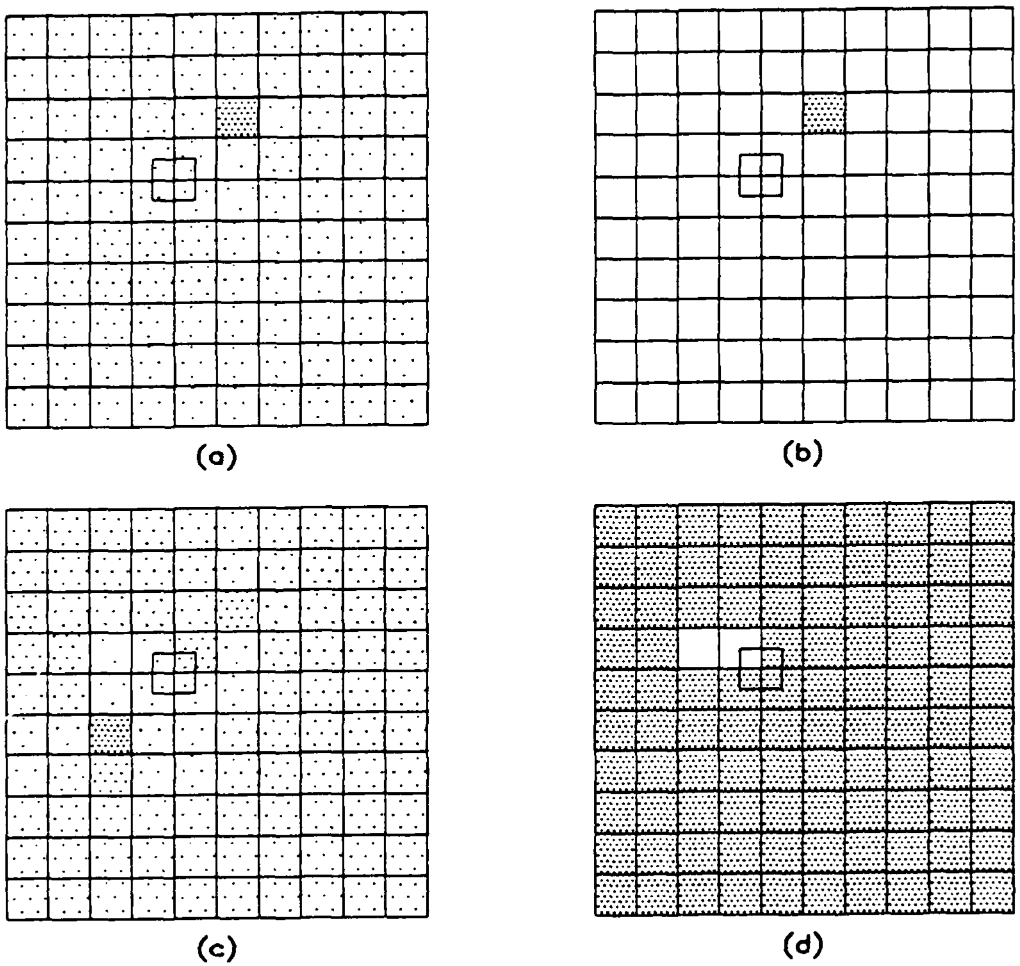


Fig. 7.38. Ray tracing results for high permittivity cylinder using TOF measurements. (a) Filtered reconstructed image. (b) MMP of image in (a). (c) Filtered ray traced image. (d) MMP of image in (c).

which are largely refractive in nature. The ray tracing method is most effective on reconstructions where the anomaly nearly coincides with the reconstruction pixels, although methods to alleviate this dependency have been presented. In addition, care must be taken in applying this technique in that improvements may not be obtained in all cases. However, for some profiles the improvements in the reconstructed image are worth using this method.

7.6.3 Ray Optics Model Matching Method

A. Introduction

As discussed in Chapter II, and as shown in the block diagram of Fig. 7.1, it is possible to use the ray optics method (Section 2.4) to improve the reconstructed image. This process is illustrated by the block diagram in Fig. 3.3. As noted above, this process will be most effective when the profile being reconstructed contains high contrast anomalies, making diffraction the dominant effect in the measurement process. It will be demonstrated on the cross section containing the high conductivity square cylinder (profile c of Table 7.1).

From the results of combining the CW and TOF reconstructions as presented in Fig. 7.10, we have reason to suspect the presence of a high conductivity anomaly in the region of the shaded pixels in the image, (c). Therefore, we will search for the exact position of the anomaly in the four pixel region fc ed by the shaded pixels and their two mutual neighbors. Of course, the search region could be expanded at the expense of greater computation time. The search method to be used will be as follows:

a) Choose a location inside the search region.

b) Calculate the CW measurement data using the ray optics method assuming a small square cylinder centered at the chosen location.

c) Increase the size of the cylinder and calculate new CW data.

d) If the calculated CW data using the current size of cylinder more closely matches the actual measurement data, save this size for the cylinder.

e) Repeat steps c and d until the cylinder size giving the smallest error is found.

f) Choose a new location inside the search region.

g) Repeat steps b through f until the location and size of the cylinder giving the smallest error is found.

For the example to be presented, the smallest cylinder size is chosen to be 1.0 m on a side, and the size increment will be 0.5 m. The cylinder will be moved over the search region in steps of 0.25 m. As in the examples of Section 7.4, the SNR wil be equal to 30 dB.

B. Example from Test Profiles

When the search method described above was applied to the CW data for the square anomaly, profile (c) of Table 7.1, the position of minimum error was found to 7.5 m in the horizontal direction and 8.0 m in the vertical direction. The actual location of the cylinder was 8.0 m in the horizontal direction and 8.0 m in the vertical direction. Therefore, the error in position was 0.5 m. The cylinder size having smallest error was 2 m. This is the same as the size of the cylinder used to simulate the data.

Fig. 7.39 is a graphical depiction of the process of finding the cylinder location having minimum error. The graphs in this figure were generated by fixing the horizontal (vertical) position of the cylinder at the actual position and then plotting the error versus vertical (horizontal) position of the cylinder. Note that the 'vertical error' (obtained by changing the vertical position of the cylinder) has a better defined minimum than the 'horizontal error'. In addition, the 'vertical error', unlike the 'horizontal error' is monotonic on either

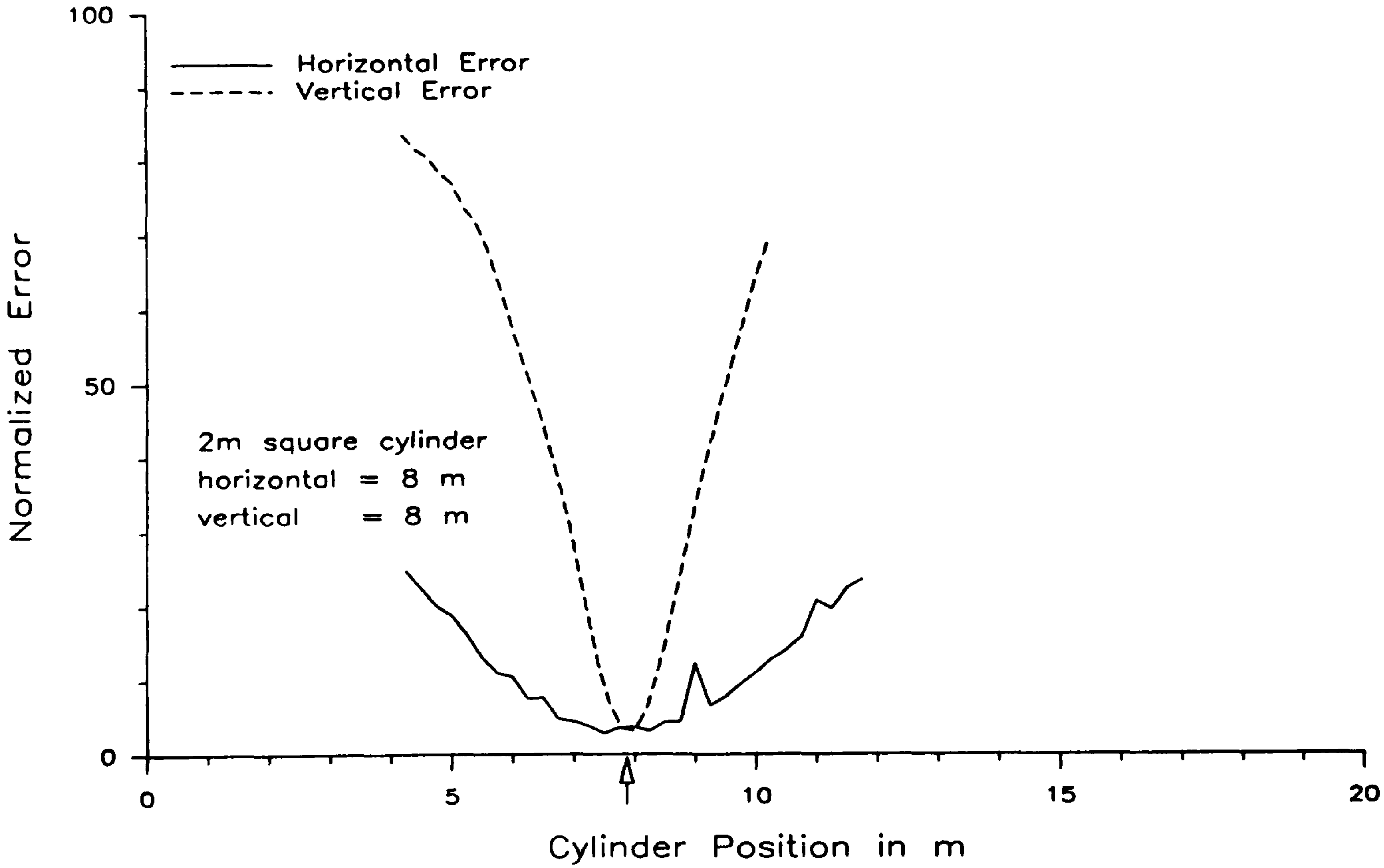


Fig. 7.39. Illustration of using the ray optics method to find the position of a square anomaly in a search region.

side of the minimum. These results are not surprising in that resolution in the vertical direction should be greater than the horizontal direction because of the scanning geometry. Note the limited view angles listed in Table 7.1. The view angles determine the horizontal resolution, while the transmitter and receiver spacings determine the vertical resolution. The transmitter/receiver spacings were sufficiently close to give good vertical resolution. In fact, this effect explains the spreading of the location of the anomaly in the horizontal direction, which was seen in the reconstructions presented at the beginning of this chapter.

C. Sample Reconstructions for Complex Anomalies

In this section we would like to demonstrate that the methods previously used are not restricted to cross sections containing only square or circular cylinders. Since simulations for arbitrarily shaped cylinders can only be obtained using the VCM (which gives CW data), the reconstructions to be presented will only be for CW data. For all of the examples, the WLS method with the CG-GPM algorithm will be used. The solution will be constrained at one third the attenuation of the background medium.

The first example features a tunnel with an arched roof located in a background medium having a conductivity of 0.001 S/m and a relative permittivity of 10. The floor of the tunnel is approximately 1.2 m across, and the height of the tunnel is also approximately 1.2 m. The reconstruction region is 20 m by 20 m, and the increment between transmitter (receiver) locations is 1 m. A total of 400 measurements of the received power will be simulated. The SNR will be equal to 30 dB.

After performing the reconstruction process as previously described, a high attenuation region 4 m in height and 8 m in width, centered around the tunnel, was identified. This region is labeled the

search region in Fig. 7.40. Also shown in the figure is the actual location of the tunnel. Since a diffraction theory does not presently exist for objects having lower index of refraction than the surrounding medium, the ray optics method for locating the tunnel cannot strictly be used. However, by noting the similarity between the magnitude responses of the tunnel and conducting cylinder in Fig. 6.1, we can attempt to locate the tunnel by using a high conducting cylinder in the search process. By using the high conducting cylinder, the ray optics method can be used to efficiently find the location of minimum error (refer to Fig. 3.3). When this operation was performed, the tunnel was 'located' 1.2 m from the actual tunnel position as shown in Fig. 7.40. This gives a much more accurate indication of the tunnel location than the raw, reconstructed image.

The second example has the same conditions as the one above, except that the anomaly has a conductivity of 0.05 S/m and a relative permittivity of 20. The cross section of the anomaly is in the shape of an 'L' as shown in Fig. 7.41. The high attenuation region in the reconstructed image is again 4 m by 8 m, as can be seen in Fig. 7.41. In this case, the process of accurately locating the anomaly using the ray optics method is very effective as the located anomaly nearly coincides with the actual anomaly.

7.7 Conclusions

In this chapter we have presented a process for reconstructing underground images and detecting and identifying high contrast anomalies in the image. This process uses both CW and TOF measurement data in order to characterize the anomalous region. The effectiveness of this process was demonstrated on a number of test cases using three different algorithms. Of the three algorithms, the CG-GPM gave the best reconstructions.

The ability of the WLS method (used in finding the reconstructed

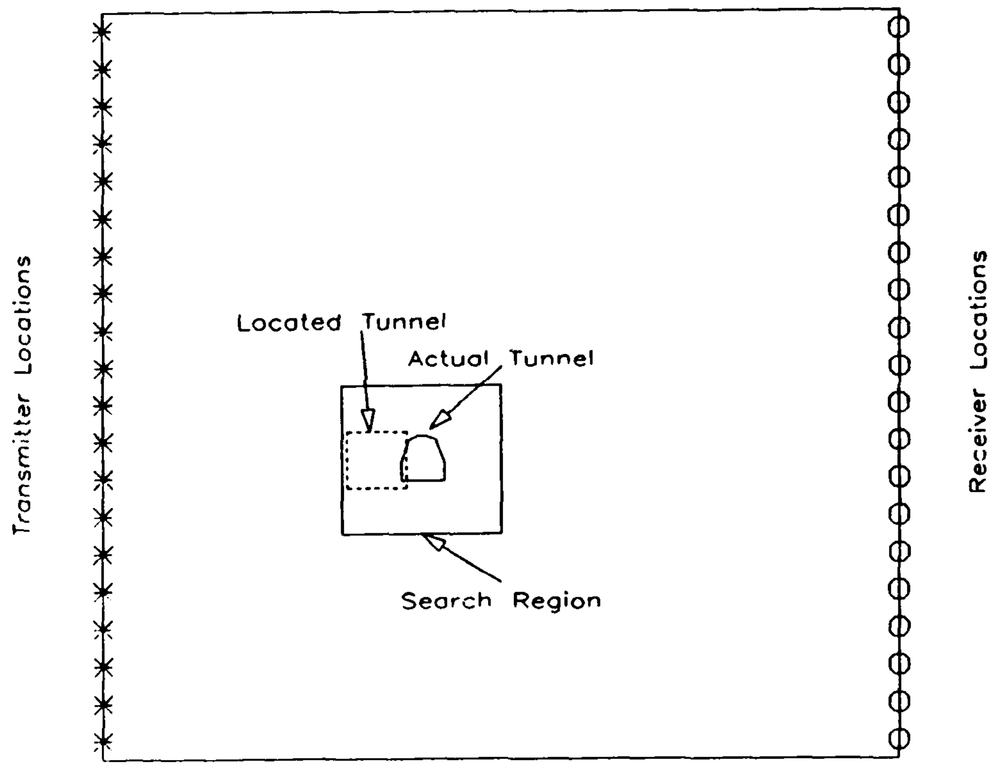


Fig. 7.40. Reconstruction showing ray optics refinement for an arched tunnel in a homogeneous earth.

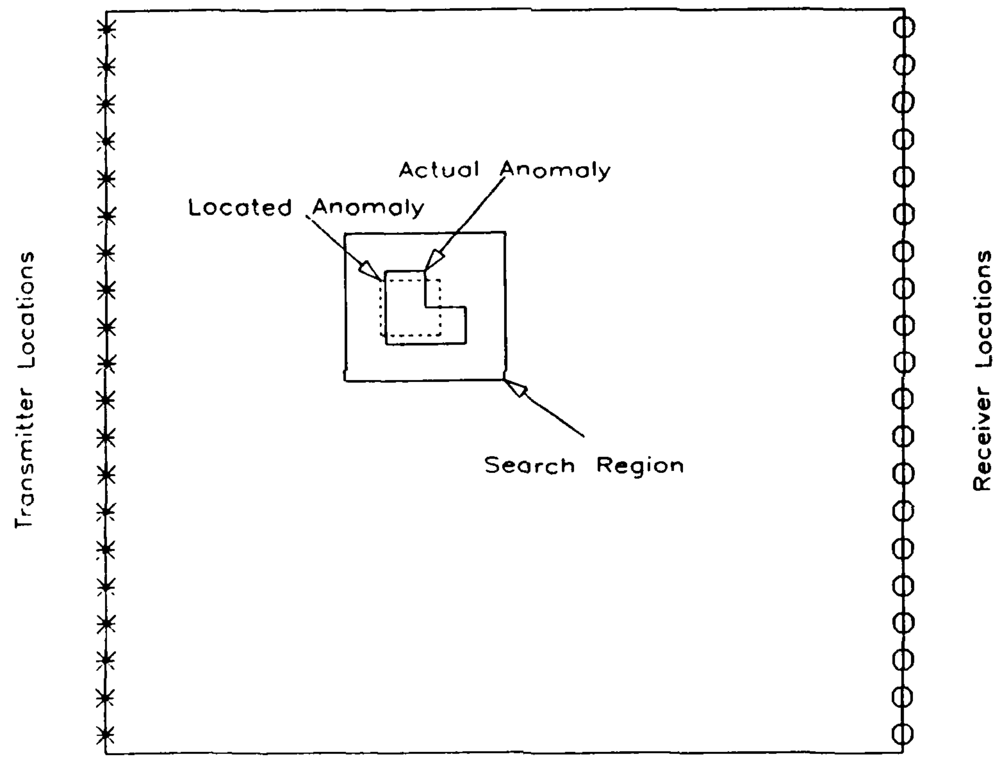


Fig. 7.41. Reconstruction showing ray optics refinement for an 'L' shaped anomaly in a homogeneous earth.

image) to diminish the effects of additive noise was also demonstrated. In all cases, using this method resulted in better images. The application of constraints on the solution image was shown to reduce artifacts due to diffraction phenomena. This reduction was evident for two test cases and for different noise levels.

Of importance to the reconstruction process is the ability to accurately locate and determine the size of an anomaly. This would be important if, for example, it were required to accurately drill into an underground stream or a tunnel. Since the electromagnetic rays do not follow straight paths, the size, shape, and location of the anomalous region in the reconstructed image does not match the actual anomaly. For this reason, we have used the ray optics model as a tool in refining the size and location of the anomaly. This takes the form of the ray tracing method for cross sections in which refraction is the dominant effect and takes the form of the model matching method when diffraction is the dominant effect.

CHAPTER VIII

SUMMARY AND CONCLUSIONS

8.1 Summary of Results

In this report some new methods for geotomography are presented. As indicated by the block diagram in Fig. 7.1, geotomography is not a one step process, and our goal has been to make improvements to all of the steps illustrated in the block diagram. In addition, in order to obtain good reconstructions, one must use as much *a priori* information as possible. This is accomplished through the use of the weighted least squares method and the application of constraints to the reconstruction process.

As a first step in understanding and characterizing the image reconstruction problem, it is first necessary to develop good models of the process. Chapter II reviews some of the models used for the cross-borehole geometry. A feature lacking in these models is the ability to characterize the diffraction of electromagnetic rays from objects located in the earth. Note that some of the models implicitly include diffraction effects, but give no indication of the relative magnitude of the contribution of these effects to the total response. Therefore, ray diffraction theory has been adapted to this application. In this way, the ray optics model is able to explicitly predict diffraction phenomena, and quantify its effects.

In Chapter III some of the standard reconstruction methods for geophysical applications are reviewed. The algebraic method, which uses a straight ray model, is chosen since it is the most robust method. This is in line with our focus on locating, detecting, and identifying high contrast anomalies. The major shortcoming with the algebraic method is that it ignores diffractions, refractions, and reflections. It is for this reason that a new method (i.e., the model matching

method) based on the ray optics model of Chapter II has been introduced. This method can be used to verify and refine a reconstructed image obtained using the algebraic method. In addition, Chapter III discusses the general theory of inverse problems. Regularization methods are presented to reduce the effects of noise in the measurement data.

One means of regularization is to find an approximate solution using least squares techniques. These techniques are investigated in Chapter IV. An extension of the least squares method is the weighted least squares method. This method allows the incorporation of *a priori* knowledge into the solution. Specific approaches for achieving this goal are presented.

The singular value decomposition (SVD) is a powerful tool for examining the behavior of least squares algorithms. This decomposition is exploited for developing the algorithms in Chapter IV. In addition, this decomposition leads to an algorithm which is able to incorporate constraints into the solution. Unfortunately, the SVD algorithm is not useful for reconstructing large images. In these cases an iterative algorithm is needed. The conjugate gradient (CG) is one such algorithm having many desireable properties. However, this algorithm can not directly incorporate constraints. Therefore, a method using the CG algorithm with the gradient projection method is developed. In this way an explicit method for finding a constrained solution is obtained.

Chapter V discusses some methods of processing raw, reconstructed images in order to reduce noise artifacts and detect subsurface anomalies. A new method of detecting high contrast anomalies is introduced, and its superiority over a standard technique is demonstrated.

In Chapter VI a discussion of the importance of obtaining both continuous wave (CW) and time-of-flight (TOF) measurements when scanning between two boreholes is given. It is important to obtain both types of measurements since:

a) Using one of the measurement processes alone does not uniquely

determine a reconstructed image.

b) The two measurement processes are susceptible to different types of errors. Therefore, both should be used to maximize the probability of detecting an anomaly.

A viable method of incorporating both types of measurements into the reconstruction process is suggested.

Chapter VII summarizes the results of the preceding chapters into a process for reconstructing a cross sectional image and then detecting, locating, and identifying anomalies in the region. The efficacy of this process is demonstrated on some geophysical sample cross sections. Once an anomalous region is detected, ray tracing or the model matching procedure can then be used to pinpoint the location of the anomaly, to identify it, and to estimate its size. These procedures are demonstrated on a number of sample cross sections.

8.2 Conclusions

The image reconstruction problem has been shown to be very ill-posed in the geotomography setting. Accordingly, algorithms are presented which are numerically stable and are able to incorporate *a priori* information into the reconstruction process. Of these algorithms, the gradient projection method using conjugate gradient minimization (CG-GPM) is found to be the best algorithm for incorporating inequality constraints into the solution. In addition, this algorithm achieves fast convergence. The weighted least squares (WLS) method should be used in conjunction with the CG-GPM in order to reduce the effects of noise in the measured data.

The algorithms that were developed can be used on data generated using continuous wave (CW) or time-of-flight (TOF) subsurface measurements. To obtain maximum information about an underground cross section being scanned, it is recommended that both types of measurements should be made in the reconstruction process.

The reconstructed underground image can be improved using ray tracing or the model matching procedure. If it is known *a priori* that the subsurface anomalies will cause refraction to be the dominant effect, then ray tracing should be used. If diffraction effects are dominant, then the model matching procedure is the best choice. If no knowledge about the subsurface anomalies is available, then both methods should be attempted to determine which one leads to improvements in the reconstructed image.

REFERENCES

[1] R. N. Bracewell, "Radio interferometry of discrete sources," *Proc. IRE,* vol. 46, no. 106, pp. 97-105, Jan. 1958.

[2] W. H. Olendorf, "Isolated flying-spot detection of radio density discontinuities; displaying the internal structures pattern of a complex objects," *IRE Trans. Bio-med. Electron.,* BME-8, pp. 68-72, 1961.

[3] A. M. Cormack, "Representation of a function by its line integrals, with some radiological applications," *J. Appl. Phys.,* 34, pp. 2722-2727, 1963.

[4] G. N. Hounsfield, "Computerized transverse axial scanning tomography: Part I, description of the system," *Br. J. Radiol.,* 46, pp. 1016-1022, 1973.

[5] J. Radon , "Uber die Bestimmung von Funktionen durch Integralwerte langs gewisser Mannigfaltigkeiten," Berichte Sachsische Akademie der Wissenschaften. Leipzig, Math. Phys. Kl., 69, pp. 262-267, 1917.

[6] J. Hadamard, *Lectures on the Cauchy Problem in Partial Differential Equations*. New Haven: Yale University Press, 1923.

[7] J. P. Barton, "Feasability of neutron radiography for large bundles of fast reactor fuel," *Tech. Rept. IRT 6247-004,* Instrumentation Research Technology Corporation, San Diego, CA, 1978.

[8] G. Nolet, Ed., *Seismic Tomography with Applications in Global Seismology and Exploration Geophysics*. Dordrecht: D. Reidel Publishing Co., 1987.

[9] J. J. Holmes and C. A. Balanis, "Electromagnetic modeling of reflections inside an in situ gasified coal seam," *Radio Sci.,* vol. 12, no. 1, pp. 33-40, Jan. 1977.

[10] E. F. Laine, "Remote monitoring of the steam-flood enhanced oil recovery process," *Geophysics*, vol. 52, no. 11, pp. 1457-1465, Nov. 1987.

[11] R. J. Lytle, E. F. Laine, D.L. Lager, and D.T. Davis, "Cross-borehole electromagnetic probing to locate high-contrast anomalies," *Geophysics*, vol. 44, no. 10, pp. 1667-1676, Oct. 1979.

[12] D .L. Lager and R. J. Lytle, "Determining a subsurface electromagnetic profile from high-frequency measurements by applying reconstruction technique algorithms," *Radio Sci.*, vol. 12, no. 2, pp. 249-260, 1972.

[13] K. A. Dines and R. J. Lytle, "Computerized geophysical tomography," *Proc. IEEE*, vol. 67, no. 7, pp. 1065-1073, July 1979.

[14] R. J. Lytle, K. A. Dines, "Iterative ray tracing between boreholes for underground image reconstruction," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-18, no. 3, pp. 234-240, July 1980.

[15] A. H. Andersen and A. C. Kak, "Digital ray tracing in two-dimensional refractive fields," *J. Accoust. Soc. Am.*, vol. 72, no. 5, pp. 1593-1606, Nov. 1982.

[16] R. D. Radcliff and C. A. Balanis, "Electromagnetic Geophysical incorporating Refraction and Reflection," *IEEE Trans. APS*, vol. AP-29, no. 2, pp. 288-292.

[17] R. D. Radcliff and C. A. Balanis, "Reconstruction algorithms for geophysical applications in noisy environments," *Proc. IEEE*, vol. 67, no. 7, pp. 1060-1064, July 1979.

[18] R. D. Radcliff, *Electromagnetic Geophysical Tomograpy Incorporating Refraction and Reflection*, Ph.D. dissertation, West Virginia Univ., 1984.

[19] A. J. Devaney, "Geophysical diffraction tomography," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-22, no. 1, pp. 3-13, Jan. 1984.

[20] M. Slaney, A. C. Kak, and L. E. Larsen, "Limitations of imaging with first-order diffraction tomography," *IEEE Trans. MTT*, vol. MTT-32, no. 8, pp. 860-874, Aug. 1984.

[21] M. S. Frank and C. A. Balanis, "Geometrical optics and diffraction techniques for data simulation in cross-borehole tomography," *Presented at IGARSS '88,* Edinburgh, Scotlaind.

[22] C. A. Balanis and J. D. Bentley, "Algorithm and filter selection in geophysical tomography," *IEEE Trans. Geosci. Remote Sensing,* vol. GE-24, no. 6, pp. 983-996, Nov. 1986.

[23] B. L. Carlson and C. A. Balanis, "Pulse time-delay method for geophysical tomography," *IEEE Trans. Geosci. Remote Sensing,* vol. GE-26, No. 5, pp. 511-517, Sept. 1988.

[24] G. R.Olhoeft, "Electrical properties from 10^{-3} to 10^{9} Hz -- Physics and chemistry --," *Proc. of 2nd Intl Symposium on the Physics and Chemistry of Porous Media,* Schlumberger-Doll, Ridgefield, CT, Am. Inst. Phys., Conf. Proc. 154, pp. 281-298, Oct. 1986.

[25] R. N. Grubb, P. L. Orswell, and J. H. Taylor, "Borehole measurements of conductivity and dielectric constant in the 300 kHz to 25 MHz frequency range," *Radio Science,* vol. 11, no. 4, pp. 275-284, April 1976.

[26] J. R. Wait, *Electromagnetic Waves in Stratified Media*. Oxford: Pergammon Press, 1962.

[27] K. B. Wolf, *Integral Transforms in Science and Engineering*. New York: Plenum Press, pp. 365-368, 1979.

[28] M. L. Oristaglio and G. W. Hohmann, "Diffusion of electromagnetic fields into a two-dimensional earth: a finite-difference approach," *Geophysics,* vol. 49, no. 7, pp. 870-894, July 1984.

[29] A. Erdelyi, et al., *Tables of Integral Transforms*. vol. 2, New York: McGraw-Hill, pp. 8-11, 1954.

[30] R. F. Harrington, *Time-Harmonic Electromagnetic Fields*. New York: McGraw-Hill Book Company, 1961.

[31] A.Q. Howard, *Mode Match Solutions to Electromagnetic Scattering Problems in a Conducting Half-Space Environment,* Ph.D. Thesis, University of Colorado, 1972.

[32] N. N. Lebedev, *Special Functions and Their Applications*. Englewood Cliffs, NJ: Prentice-Hall, pp. 128-131, 1965.

[33] D. E. Amos, "A portable package for Bessel functions of a complex argument and nonnegative order," *ACM Trans. Math. Soft.*, Sept. 1986.

[34] J. H. Richmond, "Scattering by a dielectric cylinder of arbitrary cross section," *IEEE Trans. Antennas Propagation*, vol. AP-13, pp. 334-341, May 1965.

[35] A. Q. Howard and J. L. Kretzschmar, "Synthesis of EM geophysical tomographic data," *Proc. IEEE*, vol. 74, no. 2, pp. 353-360, Feb. 1986.

[36] R. F. Harrington, *Field Computation by Moment Method*. New York: McGraw-Hill, pp. 223-232, 1961.

[37] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. New York: IEEE Press, pp. 204-218, 1988.

[38] T. E. Owen, "Cavity detection using VHF hole-to-hole electromagnetic techniques," *Presented at the Symposium on Tunnel Detection*, Golden, CO, July 21-23, 1986.

[39] R. D. Radcliff and C. A. Balanis, "Modified propagation constants for nonuniform plane wave transmission through conducting media," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-20, no. 3, pp. 408-411, July 1982.

[40] J. B. Keller, "Geometrical theory of diffraction," *J. Opt. Soc. Amer.*, vol. 52, pp. 116-130, Feb. 1962.

[41 R. G. Kouyoumjian and P. H. Pathak, "A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface," *Proc. IEEE*, vol. 62, no. 11, pp. 1448-1461, Nov. 1974.

[42] T. Senior, "Impedance boundary conditions for imperfectly conducting surfaces," *Appl. Sci. Res.*, Sect. B, vol. 8, pp. 418-436, 1960.

[43] R. Tiberio, G. Pelosi, and G. Manara, "A uniform GTD formulation for the diffraction of a wedge with impedance faces," *IEEE Trans. Antennas Propagation*, vol. AP-33, no. 8, pp. 867-872, Aug. 1985.

[44] S. R. Deans, *The Radon Transform and some of its Applications*. New York: John Wiley & Sons, 1983.

[45] H. W. Engl and C. W. Groetsch, *Inverse and Ill-Posed Problems*. New York: Academic Press, 1987.

[46] G. T. Herman and H. K. Tuy, "Image Reconstruction from Projections: An Approach from Mathematical Analysis," in *Basic Methods of Tomography and Inverse Problems*. P.C. Sabatier, Ed., Bristol: Adam Hilger, pp. 3-126, 1987.

[47] S.W. Rowland, "Computer implementation of image reconstruction formulas," in *Image Reconstruction from Projections*. G.T. Herman, Ed., vol. 32 of Topics in Applied Physics, New York: Springer, pp. 7-79, 1979.

[48] G. T. Herman, *Image Reconstructions from Projections: The Fundamentals of Computerized Tomography*. New York: Academic Press, 1980.

[49] M. M. Ney, A. M. Smith, and S. S. Stuchly, "A Solution of Electro magnetic Imaging using Pseudoinverse Transformation," *IEEE Trans. Med. Imaging*, vol. MI-3, no. 4, pp. 155-162, Dec. 1984.

[50] A. N. Datta and B. Bandyopadhyay, "A Simplified First-Order Reconstruction Algorithm for Microwave Tomography," *Int. J. Electronics*, vol. 58, no. 5, pp. 831-838, 1985.

[51] C. B. Wason, J. L. Black, and G. A. King, "Seismic Modeling and Inversion," *Proc. IEEE*, vol. 72, no. 10, pp. 1385-1393, Oct. 1984.

[52] N. Bleistein and J. K. Cohen, "Nonuniqueness in the Inverse Source Problem in Acoustics and Electromagnetics," *J. Math. Physics*, vol. 18, no. 2, pp. 194-201, Feb. 1977.

[53] F. Natterer, *The Mathematics of Computerized Tomography*. Teubner: John Wiley & Sons, pp. 85-101, 1986.

[54] A. V. Tikhonov and V. Y. Arsenin, *Solution of Ill-Posed Problems*. Washington: Winston/Wiley, 1977.

[55] G. Talenti, Ed., *Inverse Problems*. no. 1225 Springer Lecture Notes in Mathematics, Berlin: Springer-Verlag, 1986.

[56] P. C. Sabatier, "Basic Concepts and Methods of Inverse Problems," in *Basic Methods of Tomography and Inverse Problems*. P.C. Sabatier, Ed., Bristol: Adam Hilger, pp. 471-668, 1987.

[57] A. K. Louis and F. Natterer, "Mathematical Problems of Computerized Tomography," *Proc. IEEE*, vol. 71, no. 3, pp. 379-389, March 1983.

[58] G. W. Stewart, *Introduction to Matrix Computations*. Orlando: Academic Press, 1973.

[59] C. C. Lyon, *The Solution of Ill-Posed Systems of Linear Equations in the Presence of Noise*. M.S. Thesis, University of Arizona, 1984.

[60] M. Z. Nashed, "Operator-Theoretic and Computational Approaches to Ill-Posed Problems with Applications to Antenna Theory," *IEEE Trans. Ant. Propag.*, vol. AP-29, no. 2, pp. 220-231, March 1981.

[61] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore: Johns Hopkins University Press, 1985.

[62] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*. Englewood Cliffs: Prentice-Hall, 1974.

[63] A. D. Whalen, *Detection of Signals in Noise*. New York: Academic Press, pp. 374-376, 1978.

[64] K. Tanabe, "Projection Method for Solving a Singular System of Linear Equations and its Application," *Numer. Math.*, 17, pp. 203-214, 1969.

[65] A. Bjorck, "Iterative Refinement of Linear Least Squares Solution I," *BIT 7*, pp. 257-278, 1966.

[66] R. D. Radcliff, C. A. Balanis, and H. W. Hill, "A Stable Geotomography Technique for Refractive Media," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-22, no. 6, pp. 698-703, Nov. 1984.

[67] M. N. Ney, A. M. Smith, and S. S. Stuchly, "A Solution of Electromagnetic Imaging using Pseudoinverse Transformation," *IEEE Trans. Medical Imaging*, vol. MI-3, no. 4, pp. 155-162, Dec. 1984.

[68] J. Neter, W. Wasserman, and M. H. Kutner, *Applied Linear Statistical Models*. Homewood: Irwin, pp. 394-400, 1985.

[69] G. H. Golub, M. Heath, and G. Wahba, "Generalized Cross-Validation as a Method of Choosing a good Ridge Parameter," *Technometrics*, vol. 21, no. 2, pp. 215-223, May 1979.

[70] S. Kaczmarz, "Angenaherte Auflosung von Systemen linearer Gleichungen," *Bull. Acad. Polon. Sci. Lett. A*, 35, pp. 355-357, 1937.

[71] G. T. Herman, A. Lent, and S. W. Rowland, "ART: Mathematics and Applications," *J. Theor. Biol.*, 42, pp. 1-32, 1973.

[72] Y. Censor, P. P. B. Eggermont, and D. Gordon, "Strong underrelaxation in Kaczmarz's Method for Inconsistent Systems," *Numer. Math.*, 41, pp. 83-92, 1983.

[73] Y. Censor, "Row-Action Methods for Huge and Sparse Systems of and their Applications," *Siam Review*, vol. 23, no. 4, pp. 444-466, October 1981.

[74] D. G. Luenberger, *Linear and Nonlinear Programming*. Reading: Addison-Wesley, 1984.

[75] M. R. Hestenes and E. Stiefel, "Methods of Conjugate Gradients for Solving Linear Systems," *J. Res. Nat. Bur. Stan.*, vol. 49, no. 6, pp. 409-436, 1952.

[76] C. C. Paige and M. A. Saunders, "LSQR: An Algorithm for Sparse Linear Equations and Sparse Least Squares," *ACM Trans. Math. Soft.*, vol. 8, no. 1, pp. 43-71, March 1982.

[77] M. S. Frank and C. A. Balanis, "A Conjugate Direction Method for Geophysical Inversion Problems," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-25, no. 6, pp. 691-701, Nov. 1987.

[78] R. Fletcher and C. M. Reeves, "Function Minimization by Conjugate Gradients," *Computer Journal* 7, pp. 149-154, 1964.

[79] M. J. D. Powell, "Restart Procedures for the Conjugate Gradient Method," *Math. Prog.* 12, pp. 241-254, 1977.

[80] J. B. Rosen, "The Gradient Projection Method for Nonlinear Programming. Part I. Linear Constraints," *J. Soc. Indust. Appl. Math.*, vol. 8, no. 1, pp. 181-217, March 1960.

[81] J. A. Scales, "Tomographic Inversion via the Conjugate Gradient Method," *Geophysics*, vol. 52, no. 2, pp. 179-185, Feb. 1987.

[82] D. C. C. Wang, A. H. Vagnucci, and C.C. Li, "Digital Image Enhancement: A Survey," *Computer Vision, Graphics, and Image Processing* 24, pp. 363-381, 1983.

[83] J. T. Tou and R. C. Gonzalez, *Pattern Recognition System*, Reading, MA, Addison-Wesley, 1974.

[84] L. S. Kwei, *Image Enhancement of Geophysical Tomograms*, M.S. Thesis, West Virginia University, 1984.

APPENDIX A

THE VOLUME CURRENT METHOD

This is the standard method used in this dissertation for calculating the electromagnetic response of arbitrarily shaped cylinders imbedded in the earth. The method was originally developed in [34], but the description here is taken from [35].

This method can be derived from the damped wave equation for a homogeneous earth assuming a sinusoidal source term ($e^{j\omega t}$ time convention). As in Chapter 2, the total field is equal to the sum of an incident and scattered field. The scattered field can be found using a Green's function formulation as in (2-46), which is repeated here in an equivalent form

$$E^{s}(\mathbf{x}) = -\int\int_{A} \{\gamma_a^2(\mathbf{x}')-\gamma_e^2\} E(\mathbf{x}')G(\rho)\, d\mathbf{x}' \tag{A-1}$$

where the integration is over the area of the cylinder, γ_a (γ_e) is the propagation constant of the cylinder (background) , $G(\circ)$ is the Green's function solution for the homogeneous problem,

$$\mathbf{x} := (x\ y)^{T} \tag{A-2}$$

is the position vector, and

$$\rho := |\mathbf{x}-\mathbf{x}'| \tag{A-3}$$

Since the total field is the sum of the incident and scattered fields it is given by

$$E(\mathbf{x}) = E^{i}(\mathbf{x}) - \int\int_{A} \{\gamma_a^2(\mathbf{x})-\gamma_e^2\}\ E(\mathbf{x}')\ G(|\mathbf{x}-\mathbf{x}'|)\ d\mathbf{x}'. \tag{A-4}$$

The equation given above can be solved numerically for the total

field by expanding the total field as

$$E(\mathbf{r}) = \sum_{n=1}^{N} a_n I_{A_n}(\mathbf{r}), \qquad \text{(A-5)}$$

where A_n represents a partitioning of the cylinder into N non-overlapping square patches, $I_{A_n}(\mathbf{r})$ is the indicator function defined by

$$I_{A_n}(\mathbf{r}) = \begin{cases} 1 & \text{if} \quad \mathbf{r} \in A_n \\ 0 & \text{if} \quad \mathbf{r} \notin A_n \end{cases}, \qquad \text{(A-6)}$$

and a_n is the average value of total field in each A_n. The a_n can be determined by choosing a vector $\mathbf{r}_n$ such that

$$a_n = E(\mathbf{r}_n), \qquad \text{(A-7)}$$

where $\mathbf{r}_n$ is the vector to a location inside the n^{th} patch.

Substituting (A-4) into the right hand side of (A-3) gives

$$E(\mathbf{r}) = E^i(\mathbf{r}) - \int\!\!\int_A (\gamma_a^2(\mathbf{r}') - \gamma_e^2) \sum_{n=1}^{N} E(\mathbf{r}_n) I_{A_n}(\mathbf{r}') G(\rho) \, d\mathbf{r}'$$

$$= E^i(\mathbf{r}) - \sum_{n=1}^{N} (\gamma_{an}^2 - \gamma_e^2) E(\mathbf{r}_n) \int\!\!\int_A I_{A_n}(\mathbf{r}') G(\rho) \, d\mathbf{r}' \qquad \text{(A-8)}$$

where the assumption was made that the propagation constant does not vary over each A_n. The integral above can be evaluated by replacing each square patch by a circular patch of radius a. Making this

replacement, the integral becomes for the n^{th} patch

$$\int\int_{A_n} G(\rho)\, d\mathbf{x}' = \int_0^{2\pi}\int_0^a G(\rho)\rho' d\rho' d\phi'$$

$$= \int_0^{2\pi}\int_0^a \frac{j}{4} H_0^{(2)}(-j\gamma_e\rho)\rho' d\rho' d\phi'$$

$$= \frac{-\pi a}{2\gamma_e} J_1(-j\gamma_e a)\, H_0^{(2)}(-j\gamma_e\rho_n), \qquad \text{(A-9)}$$

where the explicit form for the Green's function was substituted to get he second line above, and the final result was obtained by using the addition theorem for Hankel functions. The quantity ρ_n is the distance from the observation point to the n^{th} patch.

Substituting (A-8) into (A-7), the total field can be expressed as

$$E(\mathbf{x}) = E^i(\mathbf{x}) + \sum_{n=1}^{N} (\gamma_{an}^2-\gamma_e^2)E(\mathbf{x}_n)\frac{\pi a}{2\gamma_e}J_1(-j\gamma_e a)\, H_0^{(2)}(-j\gamma_e\rho_n), \qquad \text{(A-10)}$$

where ρ_n is the distance from the observation point to the n^{th} patch The total field can be found using Galerkin's method. Multiply (A-10) by $I_{A_m}(\mathbf{x})$ and integrate over all $\mathbf{x}$ to obtain

$$\Delta^2 E(\mathbf{x}_m) = \Delta^2 E^i(\mathbf{x}_m) + \sum_{n=1}^{N} (\gamma_{an}^2-\gamma_e^2)E(\mathbf{x}_n)\frac{\pi a}{2\gamma_e}J_1(-j\gamma_e a)\, k_{nm}, \qquad \text{(A-11)}$$

where Δ is the length of the side of the square patch , m ranges from 1 to N, and

$$k_{nm} = \int_A\int I_{A_m}(X')\, H_0^{(2)}(-j\gamma_e\rho_n')\, d\mathbf{X}'. \qquad (A-12)$$

Again, substituting circular patches for the square patches and using the addition theorem, this equation can be evaluated to give

$$k_{nm} = \begin{cases} \dfrac{2}{\gamma_e^2}\,[j\pi\gamma_e a H_1^{(2)}(-j\gamma_e a) - 2j], & \text{for } m = n \\ \dfrac{j2\pi a}{\gamma_e}\, J_1(-j\gamma_e a) H_0^{(2)}(-j\gamma_e\rho_{mn}), & \text{for } m \neq n \end{cases} \qquad (A-13)$$

where ρ_{mn} is the distance from the m^{th} patch to the n^{th} patch. (A-10) can be put in the form of a matrix equation by combining the total field terms as

$$L\ c = b, \qquad (A-14)$$

where

$$L = [I_N - T], \qquad (A-15)$$

$$c = \begin{bmatrix} E(\mathbf{X}_1) \\ \vdots \\ E(\mathbf{X}_N) \end{bmatrix}, \qquad (A-16)$$

and

$$b = \begin{bmatrix} E^i(\mathbf{X}_1) \\ \vdots \\ E^i(\mathbf{X}_N) \end{bmatrix}. \qquad (A-17)$$

The matrix I_N is the identity matrix in (A-15) and the elements of T

are

$$\tau_{mn} = \begin{cases} (\gamma_{am}^2-\gamma_e^2)\dfrac{\pi a}{\gamma_e^3} J_1(-j\gamma_e a)[j\pi\gamma_e a H_1^{(2)}(-j\gamma_e a) + 2j] & \text{for } m=n \\ (\gamma_{am}^2-\gamma_e^2)\dfrac{j\pi^2 a^2}{\gamma_e^2}\{J_1(-j\gamma_e a)\}^2 H_0^{(2)}(-j\gamma_e \rho_{mn}) & \text{for } m\neq n \end{cases} \quad \text{(A-18)}$$

The total field at each of the patches can be found by inverting (A-14) as

$$c = L^{-1} b. \quad \text{(A-19)}$$

and substituting this into (A-10) to find the total field at the observation point.

APPENDIX B

SOME USEFUL RESULTS FROM LINEAR ALGEBRA

In this appendix we review some results from linear algebra and the solution of linear systems of equations. The following notation will be used.

NOTATION

$Ax = b$: matrix equations arising from linearized geophysical model.

A : $m \times n$ distance matrix.

x : unknown $n \times 1$ image vector.

b : $m \times 1$ power measurement vector.

$\bar{x}$: a solution to $Ax = b$.

$\tilde{x}_{LS}$: a least squares solution to $Ax = b$.

x_k : the solution vector at the k^{TH} step of an iterative algorithm.

$r_k = Ax_k - b$: residual vector at the k^{TH} step of an iterative algorithm

p_k : an $n \times 1$ direction vector at the k^{TH} step of a conjugate direction algorithm.

$\langle x, x \rangle$: vector inner product

$\|x\|^2 = \langle x, x \rangle$: vector norm

The following notations, definitions, and formulas are necessary for an understanding of some of the ideas presented in the dissertation. All vectors will be assumed to be column vectors with real components(e.g. $x \in \mathbb{R}^n$).

B.1 Vectors

One particular vector of interest is an image vector, which is a

vector produced by taking the values of successive pixels in an image and aligning these values into a column. The norm or magnitude of a vector is the standard Euclidean norm, $\|\circ\|_2$, where the norm of an $n \times 1$ vector x with elements ξ_i will be given by:

$$\|x\|^2 = \sum_{i=1}^{n} \xi_i^{\,2} = x^T x \tag{B-1}$$

The vector inner product is defined in the usual fashion:

$$\langle x, y \rangle = \sum_{i=1}^{n} \xi_i \eta_i = x^T y \tag{B-2}$$

Two vectors are said to be orthogonal if their inner product is equal to zero. Two vectors will be A-orthogonal(or A-conjugate) if, given a matrix A:

$$\langle x, Ay \rangle = 0 \tag{B-3}$$

A functional is a mapping from a vector space to the real or complex numbers. For example, the norm of a vector is a functional. A functional of particular interest in developing conjugate direction algorithms is the quadratic functional:

$$\phi(x) = \frac{1}{2} x^T A x - x^T b \tag{B-4}$$

The span of a set of vectors is the set of all linear combinations of the given vectors, that is:

$$\text{span}(x_1\ x_2) = \text{all } y = \alpha x_1 + \beta x_2 \tag{B-5}$$

A set of vectors will be said to span a vector space if any vector in the space can be represented as a linear combination of the vectors in the set. For example, taking the vector space of all $n \times 1$ vectors, it is easily seen that any n linearly independent vectors will span this space. Two vector spaces that are needed in the sequel are the range space and null space of a matrix. The range space of a matrix A(denoted $\mathcal{R}(A)$) is the set of all vectors x such that:

$$x = Ay \tag{B-6}$$

for all vectors y. It is easy to show that this set forms a vector space. The null space of a matrix(denoted $\mathcal{N}(A)$ for the matrix A) is the set of all vectors x such that:

$$Ax = 0 \tag{B-7}$$

where 0 represents the zero vector.

B.2 Matrices

The transpose of a matrix is the matrix resulting from an interchange of the rows and columns of a matrix. A symmetric matrix is a matrix which is equal to its transpose. It is easily shown that for a general matrix, A,

$$(A^T A)^T = A^T A \tag{B-8}$$

Another property of interest is that of positive definiteness. The matrix A is positive definite if:

$$(x^T A x) > 0, \text{ for all } x \neq 0 \tag{B-9}$$

If a matrix is real and symmetric this condition translates into requiring that all eigenvalues of the matrix are greater than zero. The condition number of a matrix gives an indication of how close the matrix is to a non-invertible(singular) matrix. A matrix that is nearly singular (i.e., has a large condition number) will be susceptible to small errors in the b vector when solving the equation Ax = b. In this paper the condition number will be defined to be:

$$\kappa(A) = \frac{\lambda_{max}(A^T A)}{\lambda_{min}(A^T A)} \tag{B-10}$$

Where $\lambda_{max(min)}(\circ)$ denotes the largest(smallest) eigenvalue of the given matrix.

B.3 Solving Linear Equations

We are interested in solving the linear equation Ax = b, where in the problem at hand, x is an unknown image vector. A solution of this equation will be given by $\tilde{x} = A^{-1}b$ when the matrix is square and of full rank(i.e. it has no linearly dependent rows or columns). When the matrix A has m rows and n columns and it is not of full rank an $\tilde{x}$ cannot be found which solves Ax = b exactly. However, a least squares solution, $\tilde{x}_{LS}$, can be found such that $\|A\tilde{x}_{LS} - b\|$ is minimized(i.e. this solution 'fits' the equation in the norm sense). A least squares solution is the only feasible solution available in applications where noise and errors are introduced into the measurement and modeling processes.

B.4 Least Squares Solutions

For convenience, in the following the matrix A will be assumed to be square and of dimension n. If the matrix A is of full rank (= n), then the range space of A is the entire n-dimensional space. In this case the equation Ax=b has a unique solution as noted above. If the matrix A is of less than full rank(e.g. rank(A) = m < n), then the range space of A is an m-dimensional space which provides a natural splitting of the entire space into:

$$\mathbb{R}^n = \mathcal{R}(A) \oplus \mathcal{R}(A)^{\perp} \tag{B-11}$$

where $\mathcal{R}(A)^{\perp}$ is the orthogonal complement of the range of A (that is, every vector in $\mathcal{R}(A)^{\perp}$ is orthogonal to any vector in $\mathcal{R}(A)$). For this rank deficient case the equation Ax=b may not have a solution for all b since an arbitrary vector b can be written as:

$$b = \alpha_1 x_1 + \alpha_2 x_2 \tag{B-12}$$

where x_1 is in $\mathcal{R}(A)$ and x_2 is in $\mathcal{R}(A)^{\perp}$. A simple example may help illustrate this point.

$$\text{Let:}\quad A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 0 & 0 & 0 \end{bmatrix}$$

$$b = \begin{bmatrix} 6 \\ 15 \\ 20 \end{bmatrix} \tag{B-13}$$

For this example it is seen that no x solves Ax=b since it is impossible to generate the third element of the b vector by multiplying A by any vector x.

A least squares solution can be found even in this rank deficient case. The solution involves finding an $\bar{x}_{LS}$ which minimizes $\|A\bar{x}_{LS}-b\|_2$. For the example above it is seen that:

$$\bar{x}_{LS} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad \text{(B-14)}$$

is one such least squares solution since

$$A\bar{x}_{LS} - b = \begin{bmatrix} 0 \\ 0 \\ 20 \end{bmatrix} \qquad \text{(B-14)}$$

Of course, this solution is not unique because

$$\bar{x}_{LS} = \begin{bmatrix} 0.5 \\ 2.0 \\ 0.5 \end{bmatrix} \qquad \text{(B-15)}$$

also minimizes the norm of the residual vector.

Least squares solutions are important for a number of reasons. Two reasons are:

i) The least squares problem lends itself easily to analysis.

ii) The least squares problem arises naturally in estimation problems when the given data is normally distributed.

It should be noted, however, that in some applications(e.g. in situations with outlying data points) that a least squares solution may not be the optimum solution.

B.5 Normal Equations

It is now possible to develop the normal equations for the least

squares problem from the results above. First of all, the relation between the range space of a matrix and the null space of the transposed matrix is needed. It can be shown that the orthogonal complement of the range space of a matrix is equal to the null space of the transpose of the matrix, or in notational form:

$$\mathcal{R}(A)^{\perp} = \mathcal{N}(A^{T}) \tag{B-16}$$

From this it follows that $A^T(A\bar{x}_{LS}-b) = 0$, or in the more standard form:

$$A^T A\bar{x}_{LS} = A^T b \tag{B-17}$$

the so called normal equations. In the example above, it was seen that:

$$A\bar{x}_{LS} - b = \begin{bmatrix} 0 \\ 0 \\ 20 \end{bmatrix} \tag{B-18}$$

and this vector is in $\mathcal{R}(A)^{\perp}$. Also it is easy to see that:

$$A^T(A\bar{x}_{LS} - b) = 0 \tag{B-19}$$

for this example.

The normal equations are important when the equation solving algorithm (e.g. the conjugate gradient algorithm) requires a symmetric coefficient matrix. For this case instead of solving the original problem $Ax=b$, the equivalent set of equations $A^TAx = A^Tb$ can be solved for the least squares solution.

B.6 The Pseudo-Inverse

Perhaps the most widely used method for finding a least squares solution to a linear equation is through the use of the pseudo-inverse (also known as the generalized inverse) of the matrix A(denoted by A^{+}). This inverse has the following form when the matrix A is of full rank:

$$A^{+} = (A^{T}A)^{-1}A^{T} \tag{B-20}$$

This form is easily obtained by appealing to the normal equations given above. When the matrix A is not of full rank the pseudo-inverse can be found by using the singular value decomposition(SVD) algorithm. In either case, the least squares solution can then be found as:

$$\bar{x}_{LS} = A^{+}b \tag{B-21}$$

This method has two major drawbacks when applied to the problem of inverting geophysical data. The first drawback is that for geophysical applications the A matrix may be very large which precludes the computation and storage of A^{+}. The second drawback is that this method does not allow the application of constraints to the solution vector. For example there may be a need to find a solution vector close to $\bar{x}_{LS}$ but with all of its elements constrained to be positive.

B.7 Singular Value Decomposition (SVD)

The SVD is a standard decomposition of a linear operator. Its usefulness lies in its ability to diagonalize a matrix using orthogonal matrices. Its form for a rectangular m×n matrix is

$$A = U\ S\ V^{T} \tag{B-22}$$

where U (m×m) and V (n×n) are orthogonal matrices. S (m×n) is a

diagonal matrix whose elements are the square roots of the eigenvalues of A^TA. As mentioned above, this decomposition can be used to find the pseudo-inverse of the matrix A. This result will be derived for the regularized inverse A_γ which reduces to A^+ for $\gamma=0$, as follows

$$\begin{aligned}
A_\gamma &= (A^TA + \gamma I)^{-1}A^T \\
&= ([USV^T]^T USV^T + \gamma I)^{-1}[USV^T]^T \\
&= (VS^TSV^T + \gamma I)^{-1}VS^TU^T \\
&= [V(D+\gamma I)V^T]^{-1}VS^TU^T \\
&= V(D+\gamma I)^{-1}V^TVS^TU^T \\
&= V(D+\gamma I)^{-1}S^TU^T
\end{aligned} \tag{B-23}$$

where $D := S^TS$, and $(D+\gamma I)^{-1}$ is a diagonal matrix whose elements are of the form

$$(D+\gamma I)^{-1}_{ii} = \frac{1}{\sigma_i^2 + \gamma} \tag{B-24}$$

where σ_i is the i^{th} singular value of A.

APPENDIX C

THE LSQR ALGORITHM

In this appendix, the algorithm LSQR which was derived in [76] is described. The stopping criterion based on the effects of the singular values is used. The singular value decomposition should first be used to determine which singular values should be ignored (see chapter 4 for a discussion of this). If singular values beyond some value K are to be ignored, then set

$$\mathrm{CONLIM} = \frac{2\,\sigma_1}{(\sigma_k + \sigma_{k+1})}, \tag{C-1}$$

before using the algorithm below. In the following vectors will be denoted by lower case letters, while scalars will be denoted by lower case Greek letters (except for c_k and s_k which are also scalars). The algorithm for solving Ax=b is as follows

$$\beta_1 = \|b\|^2$$

$$u_1 = \frac{b}{\beta_1}$$

$$v_1 = Au_1$$

$$\alpha_1 = \|v_1\|$$

$$v_1 = \frac{v_1}{\alpha_1}$$

$$w_1 = v_1$$

$$\bar{\phi}_1 = \beta_1$$

$$\bar{\rho}_1 = \alpha_1$$

$$\rho_1 = (\rho_1^2 + \beta_1^2)^{1/2}$$

$$\eta_1 = \alpha_1^2$$

$$\delta_1^2 = \|\frac{w}{\rho_1}\|^2$$

For k = 1, ... , MAX_ITERATIONS

1) $u_{k+1} = Av_k - \alpha_k u_k$

2) $\beta_{k+1} = \|u_{k+1}\|^2$

3) $v_{k+1} = \frac{u_k}{\beta_{k+1}}$

4) $v_{k+1} = Au_{k+1} - \beta_{k+1} v_k$

5) $\alpha_{k+1} = \|v_{k+1}\|^2$

6) $v_{k+1} = \frac{v_{k+1}}{\alpha_{k+1}}$

7) $\rho_{k+1} = (\bar{\rho}_k^2 + \beta_{k+1}^2)^{1/2}$

8) $c_k = \frac{\bar{\rho}_k}{\rho_{k+1}}$

9) $s_k = \frac{\beta_{k+1}}{\rho_{k+1}}$

10) $\theta_{k+1} = s_k \alpha_{k+1}$

11) $\bar{\rho}_{k+1} = -c_k \alpha_{k+1}$

12) $\phi_k = c_k \bar{\phi}_k$

13) $\bar{\phi}_{k+1} = s_k \bar{\phi}_k$

14) $\eta_{k+1} = \eta_k + \alpha_{k+1}^2 + \beta_{k+1}^2$

15) $\delta_{k+1} = \delta_k + \|\frac{w}{\rho_{k+1}}\|^2$

16) If$(\eta_{k+1}\delta_{k+1} > \text{CONLIM}^2)$

STOP

17) $x_k = x_{k-1} + \frac{\phi_k}{\rho_{k+1}} w_k$

18) $w_{k+1} = v_{k+1} - \frac{\theta_{k+1}}{\rho_{k+1}} w_k$. (C-2)

In the algorithm above, η_k and δ_k denote the squares of the norms of the matrices B_k and D_k defined in [76]. For the definitions of the other quantities used above, see [76].

APPENDIX D

THE GRADIENT PROJECTION METHOD

The following algorithm is based on that given in [80]. It has been simplified to handle linear inequality constraints only, since these are the major constraints of interest in tomography. In addition, the CG algorithm has been substituted for the steepest descent method in performing the minimization over the working surface.

The gradient projection method for minimizing the functional

$$f(x) = \frac{1}{2} x^T Q x - x^T b, \tag{D-1}$$

where Q is symmetric and positive definite, subject to the constraints

$$\xi_{min} \leq \xi_i \leq \xi_{max}, \tag{D-2}$$

where ξ_i denotes an element of x for $1 \leq i \leq n$ is given by

1) Set elements of $x = \xi_{min}$

2) $r_0 = b - Qx$

3) $p_0 = 0$

4) $i = 1$

For $k = 1, \ldots$ MAX_ITERATIONS

1) $\beta_k = \dfrac{\|r_{k-1}\|^2}{\|r_k\|^2}$

2) If $(i>1)$

$p_k = r_{k-1} + \beta_k p_{k-1}$

Else

$p_k = r_{k-1}$

3) $\alpha_k = \dfrac{\|r_{k-1}\|^2}{\langle p_k, Qp_k \rangle}$

4) $d_k = \alpha_k P_k p_k$

5) If $(x_{k-1} + d_k)$ is feasible

$x_k = x_{k-1} + d_k$

Else

a) Choose $\alpha_k' < \alpha_k$ such that

$x_k = x_{k-1} + \alpha_k' P_k p_k$ is feasible

b) Zero the row of P_k corresponding

to the unfeasible component of d_k

k k-1 k k

6) $r_k = r_{k-1} - \alpha_k Q p_k$

7) If $d_k = 0$

a) Add row to P_k corresponding to

maximum element of r_k

b) Re-start by setting $i = 0$

8) $i = i + 1.$ (D-3)

In the above, P_k is the projection matrix which takes the direction vector onto the constraint surface.